

Linking Biochemical and Cellular Efficacy of the Coronavirus Main Protease

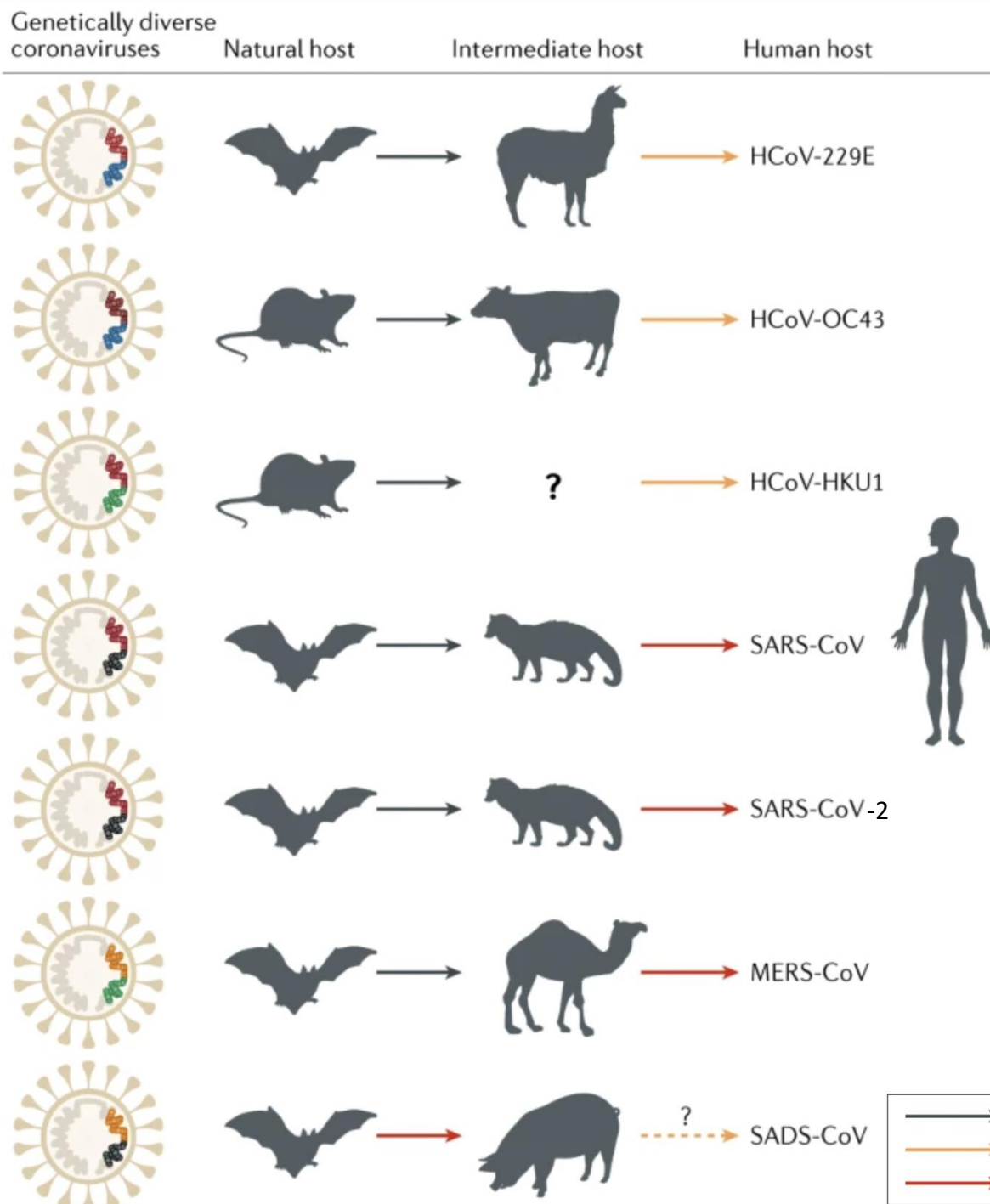
David Minh

Based on PhD thesis defense of Van La (Sophie)

in collaboration with Lulu Kang

May 14th, 2024





SARS-CoV (Severe Acute Respiratory Syndrome CoronaVirus):

- 8098 SARS infected cases and 774 deaths
=> 10% mortality [1]

MERS-CoV (Middle East Respiratory Syndrome CoronaVirus):

- 2458 infected cases and 848 deaths
=> 35% mortality [1]

SARS-CoV-2:

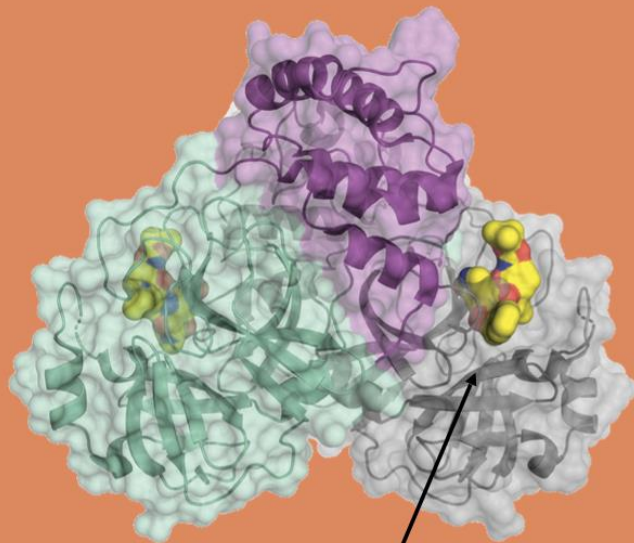
- 774,954,393 confirmed cases, 7,040,264 deaths
=> 1% mortality

[1] Ali Pormohammad *et al.*, *Rev Med Virol*, 2020

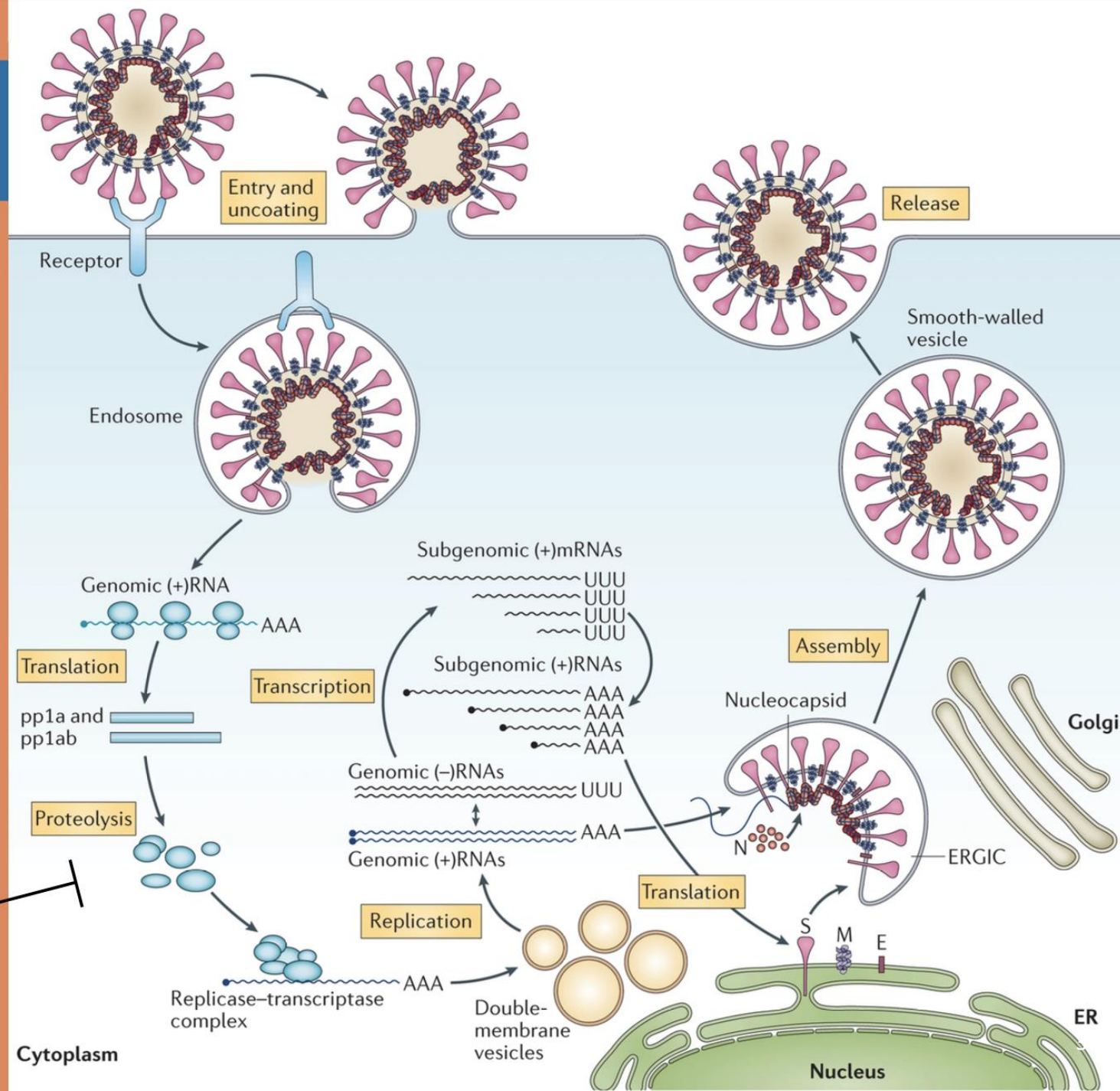
[2] <https://covid19.who.int/> on March 17, 2024

The target

3CL^{Pro}
or: M^{pro}



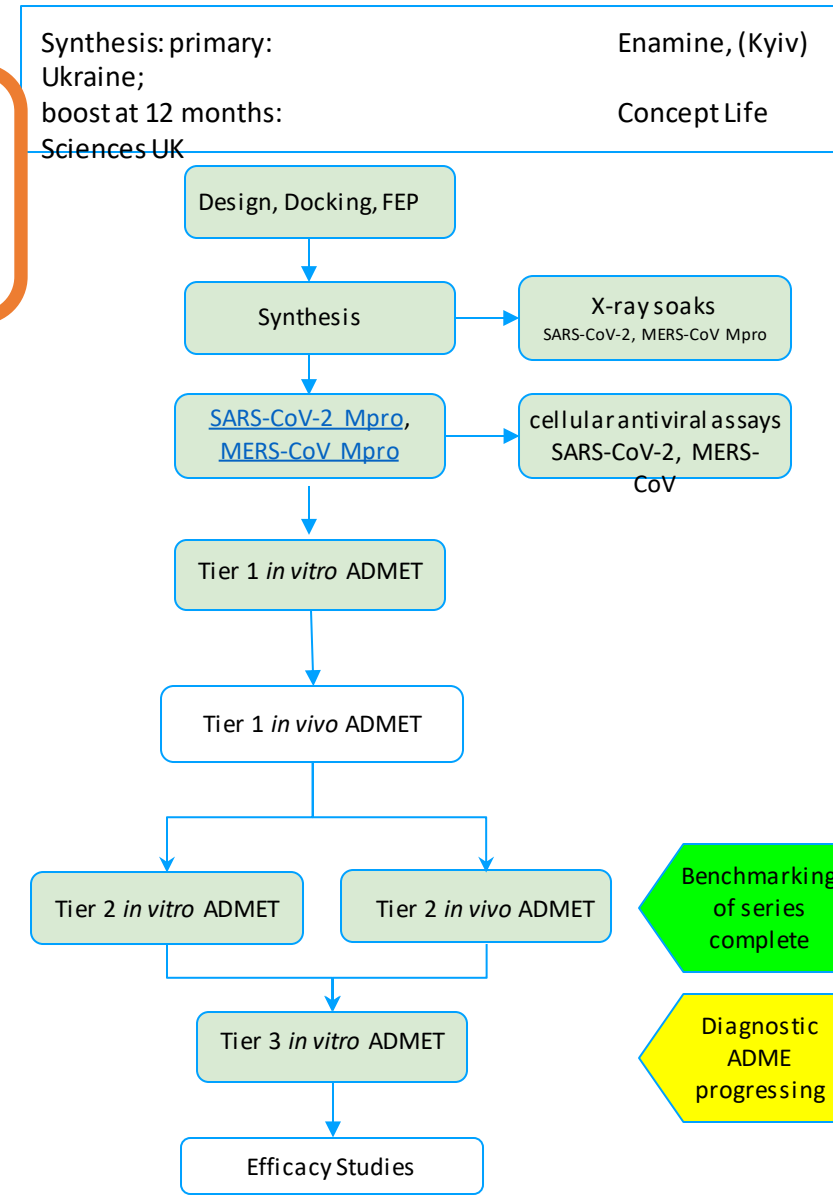
de Wit et al. Nat. Rev. Microbiology (2016)



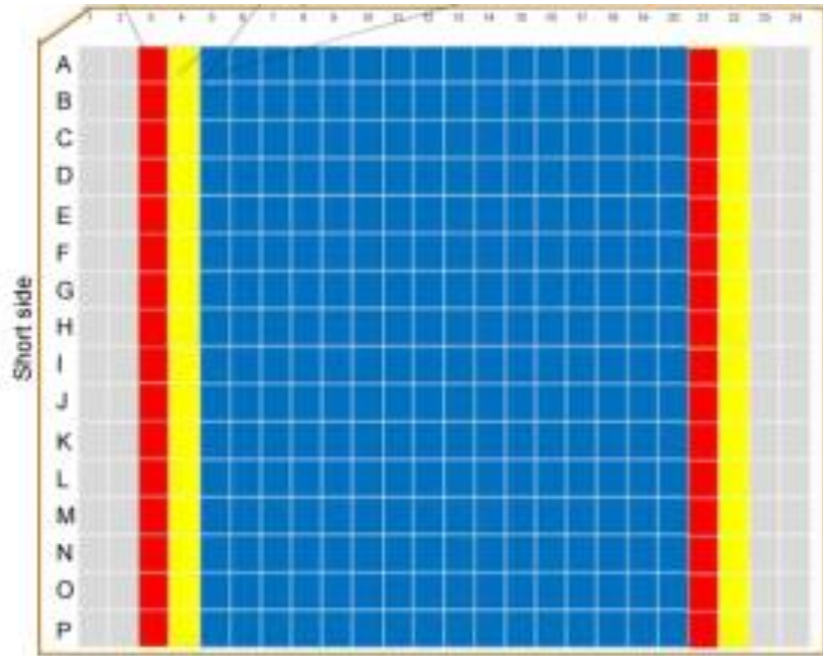
Lead Optimization

SARS-CoV-2 / MERS-CoV Mpro Assay Cascade

Area	Assays	Entry Criteria
In Vitro biochemistry (Weizmann Institute)	SARS-CoV-2 Mpro , MERS-CoV Mpro	All compounds
Cellular assay (Mount Sinai, NYC)	SARS-CoV-2, MERS-CoV	SARS-CoV-2 and MERS-CoV MPro IC50 <1uM
In vitro ADMET Tier 1 (Bienta/Enamine)	solubility (kinetic), LogD, microsome (mouse, human) MDCK Permeability (A→B)	SARS-CoV-2 and MERS-CoV MPro IC50 <1uM
In vitro ADMET Tier 2 (Concept Life Sciences, UK)	Thermodynamic Solubility, MDCK-MDR1 ER microsome (rat), hepatocyte (mouse, rat, human) Plasma Protein binding (mouse, rat, human, FCS) CYP Inhibition 5 Cyps, Patchclamp hERG, NaV 1.5, CaV 1.2	critical leads (eg for RLM: heps: bioavailability comparison) + in vivo efficacy candidates
In vitro ADMET Tier 3 (Concept Life Sciences, UK)	3A4 TDI, PXR induction, BSEP, AMES (2 Strains) Dog: Microsome, Hepatocyte, Plasma Protein – Dog As required: blood & S9 intestinal stability, transporter panels	critical leads + short list
In vivo ADMET Tier 1 (Concept Life Sciences, UK)	Mouse iv cassette 5 compounds per cassette	critical leads + rodent mics Clint < 100ml/min/mg & cellular efficacy IC50 < 1uM
In vivo ADMET Tier 2 (Concept Life Sciences, UK)	Formulation studies, Mouse iv, po bioavailability (Rat iv, po & bioavailability for human PK prediction)	in vivo efficacy candidates
Efficacy studies (Mount Sinai, NYC)	Mouse SARS-CoV-2 and Mouse MERS-CoV	expectation of free cover over EC50 for 24h
Selectivity / tox screening (Nanosyn, Eurofins)	Protease panel (61) CEREP in vitro tox panel	critical leads + short list

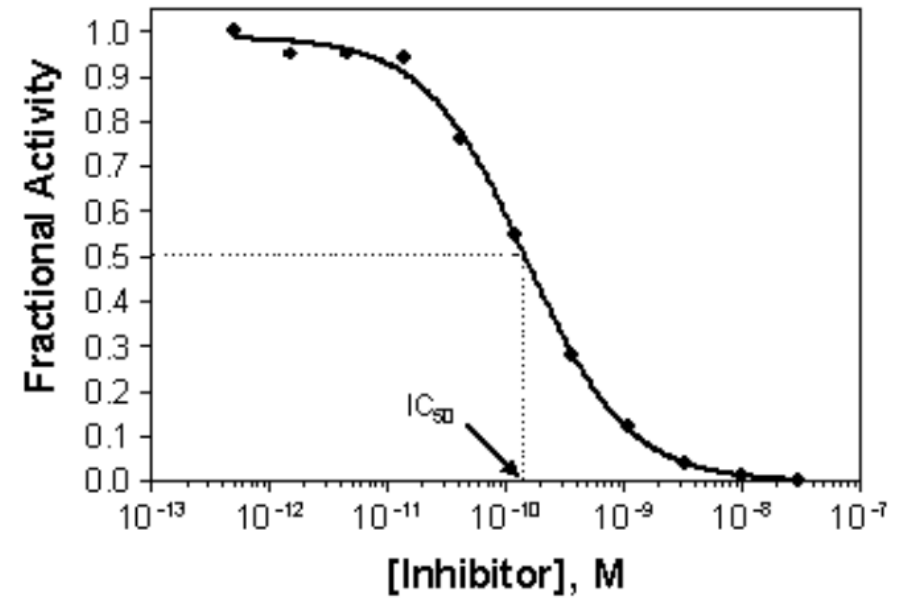


HTS: 384-well plate



Control

Concentration – response curve (CRC)



$$y_i = R_{base} + \frac{R_{max} - R_{base}}{1 + 10^{(\log Concentration_i - \log IC_{50}) * hill}}$$

(IC50: the concentration of a substance required to inhibit a specific biological response or activity by 50%. pIC50 = -logIC50)

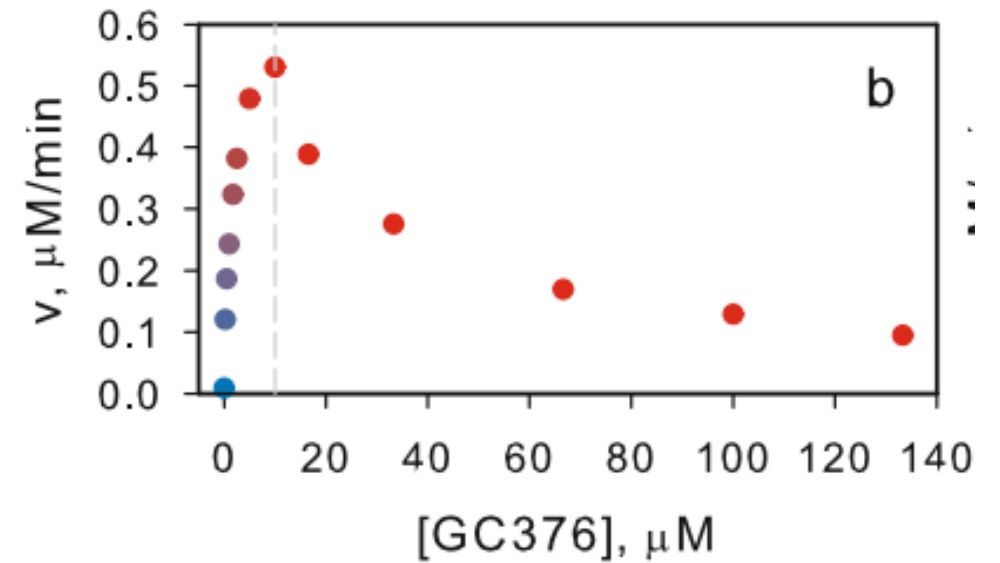
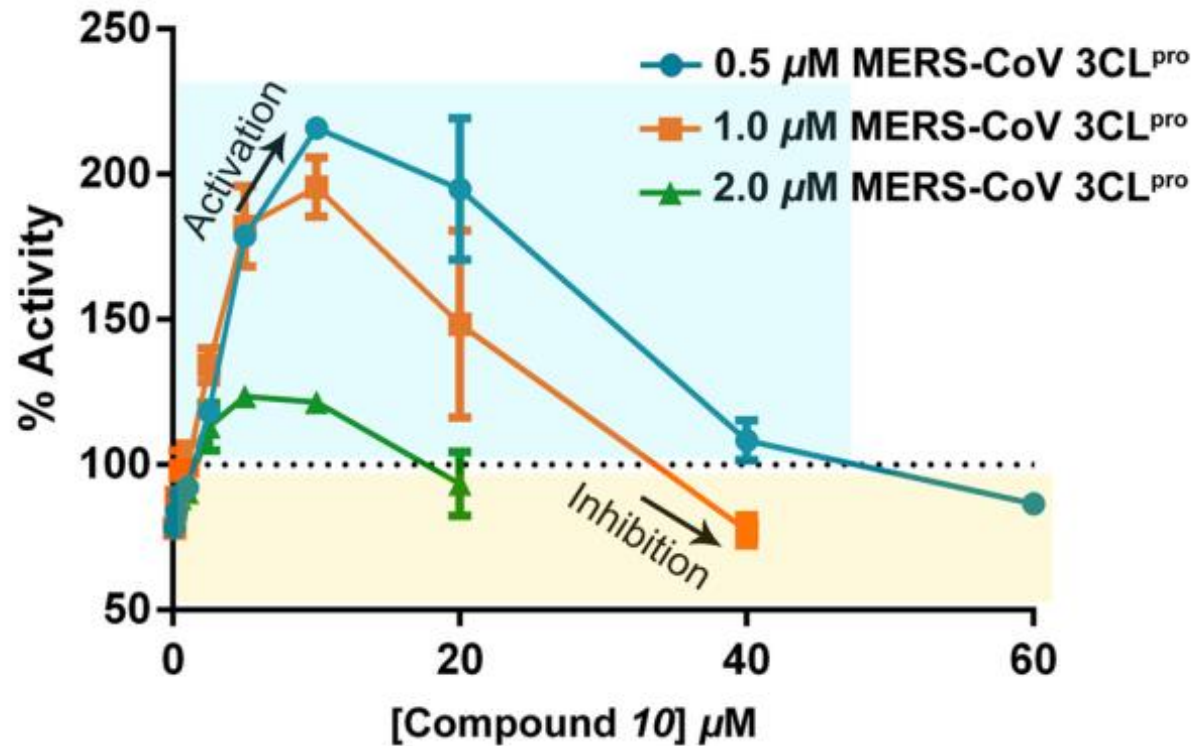
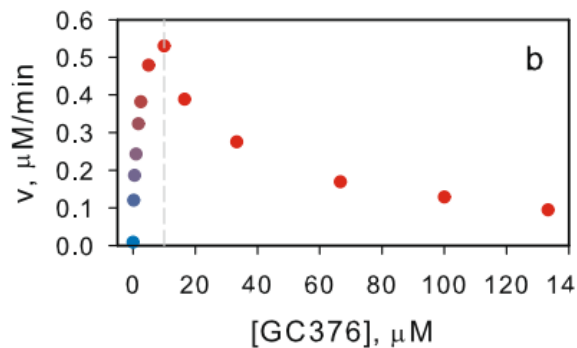


Fig. 5b. Catalytic activity of 10 μM MPro as a function of increasing GC376 concentration.

(Ref: Sakshi Tomar *et al.*, *J Biol Chem*, 2015)

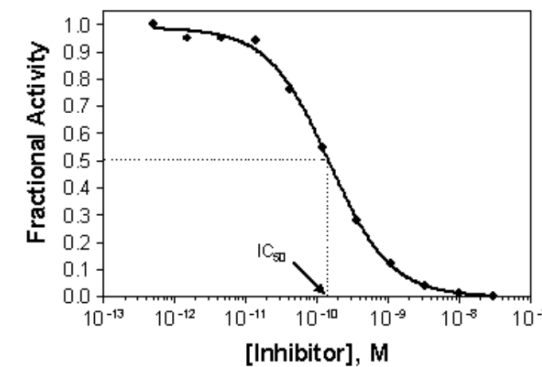
(Ref: Nashed *et al.*, *Comm Biology*, 2022)



Biochemical assay

pIC₅₀ ?

 ?

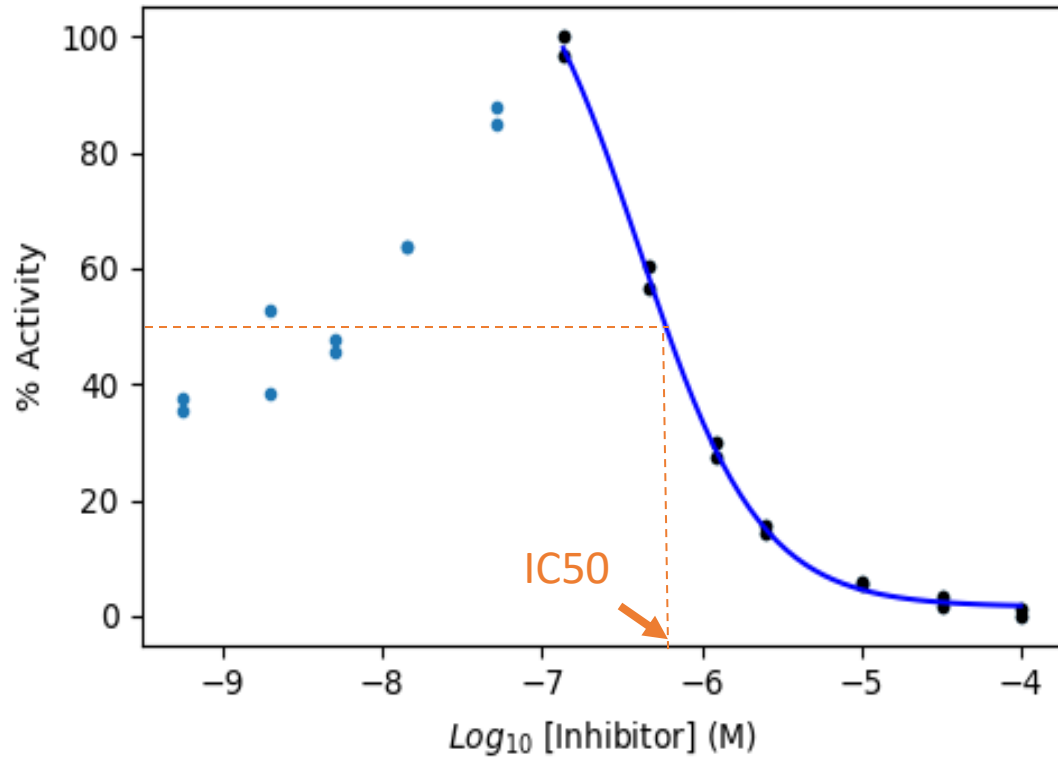


Cellular assay

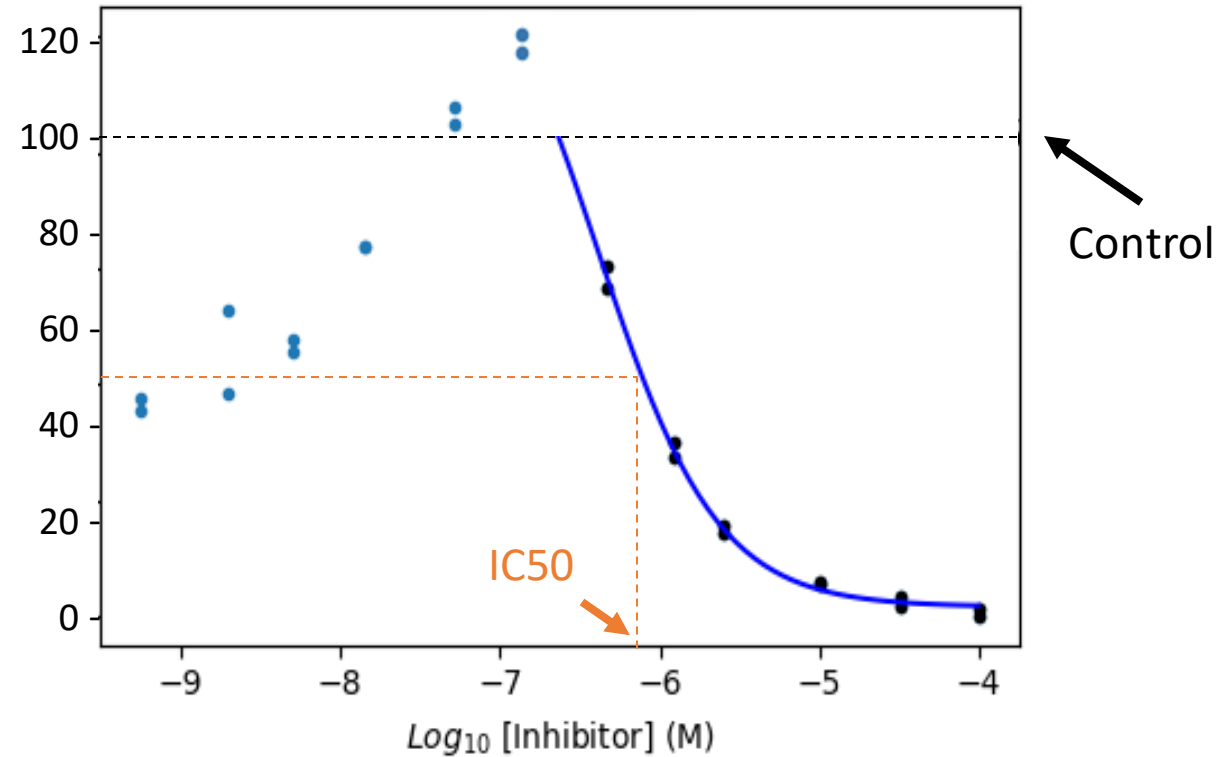
pEC₅₀

Simple procedures to fit a biphasic curve

Inhibition pIC50



Control pIC50



$$y_i = R_{base} + \frac{R_{max} - R_{base}}{1 + 10^{(\log \text{Concentration}_i - \log \text{IC}_{50}) * \text{hill}}}$$

Hypothesis

- Cellular pEC50 of MPro inhibitors is driven by the pIC50 of the MPro dimer

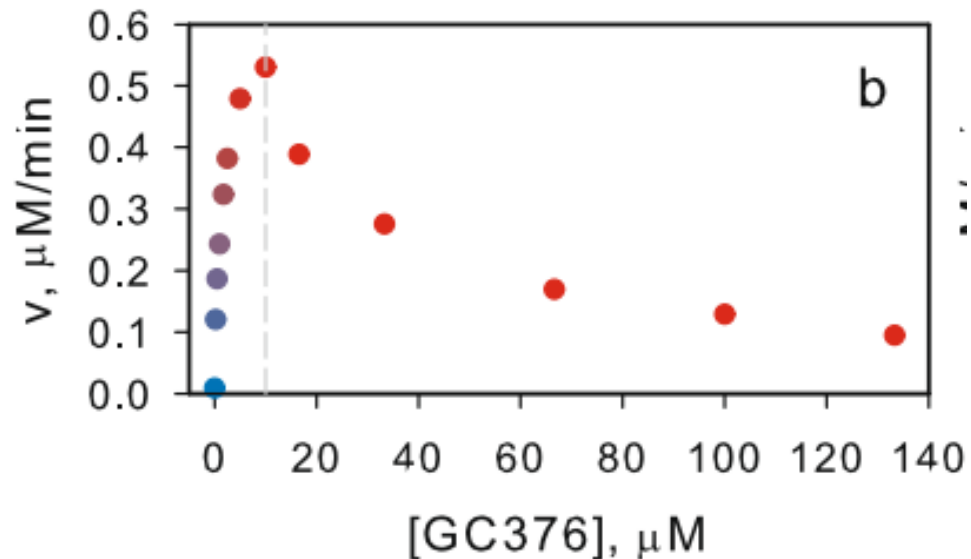
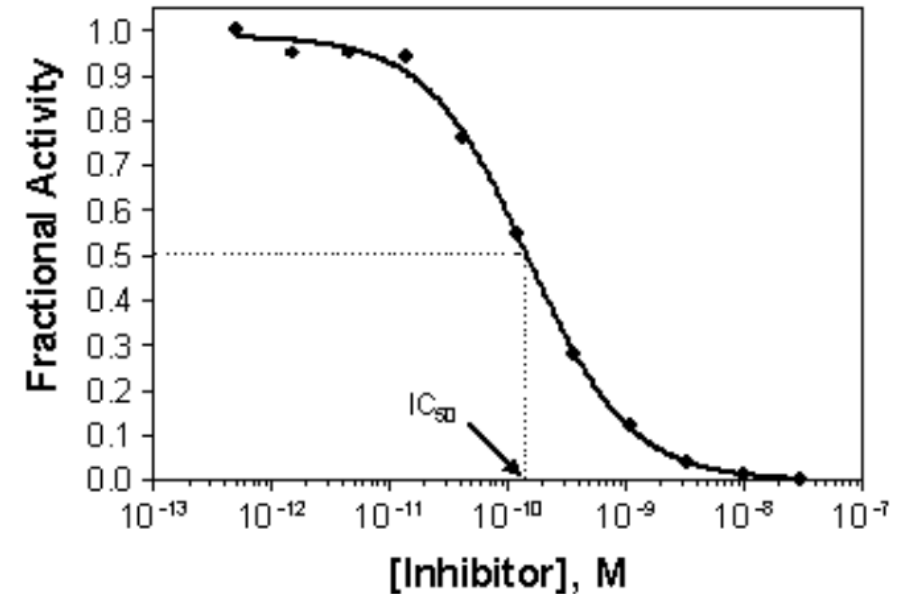
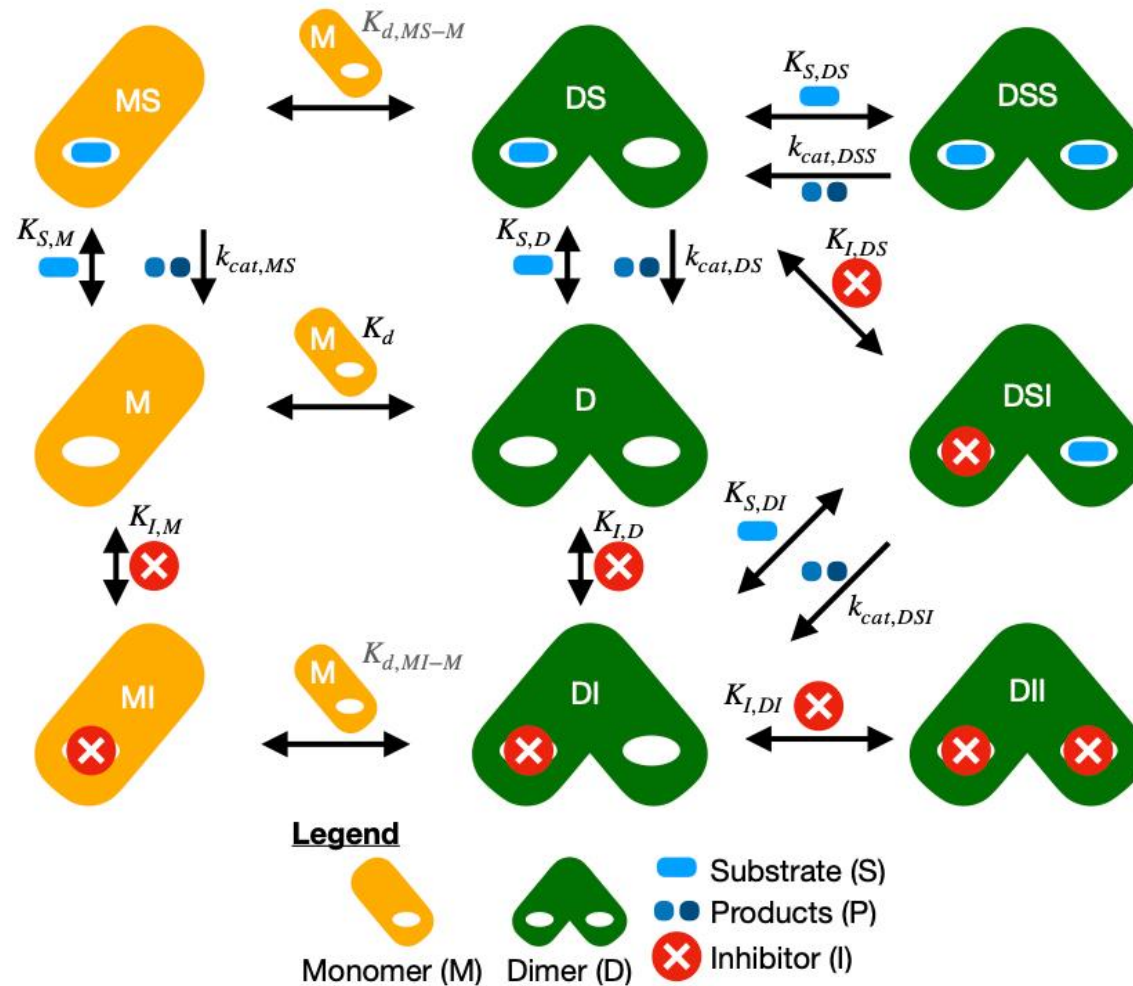


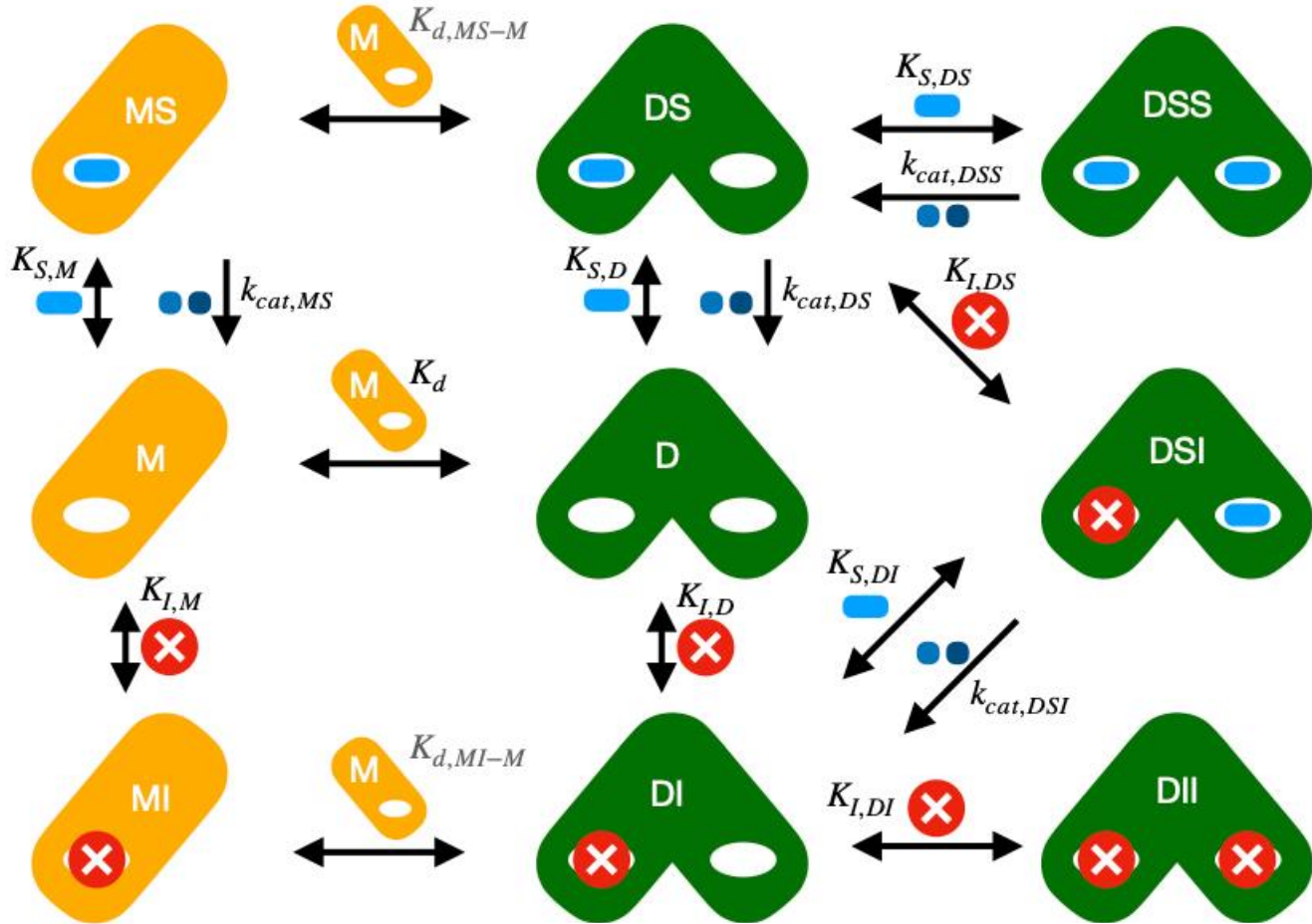
Fig. 5b. Catalytic activity of 10 μM MPro as a function of increasing GC376 concentration.



$$y_i = R_{base} + \frac{R_{max} - R_{base}}{1 + 10^{(\log Concentration_i - \log IC_{50}) * hill}}$$



Protein shown as orange rectangles for the monomer (M) or pair of overlapping green rounded rectangles for the dimer (D). Species on the top arrows are added going right/removed going left. Species to the right of arrows are added going down/removed going up. Equilibrium constants (K) are forward for the direction that leads to more complex species, with various K_d for dimerization, K_I for the inhibitor binding, and K_S for the substrate binding. Rate constants k_{cat} may depend on the dimerization and ligand binding.



$$K_d = \frac{c_M^2}{c_{DC}c^\theta} \quad K_{d,MI-M} = \frac{c_{MI}c_M}{c_{DIC}c^\theta}$$

$$K_{d,MS-M} = \frac{c_{MS}c_M}{c_{DSC}c^\theta} \quad K_{I,M} = \frac{c_{MI}c_I}{c_{MIC}c^\theta}$$

$$K_{S,M} = \frac{c_{MSC}c_M}{c_{MCS}c^\theta} \quad K_{I,D} = \frac{c_{DI}c_I}{c_{DIC}c^\theta}$$

$$K_{S,D} = \frac{c_{DSC}c_S}{c_{DSCS}c^\theta} \quad K_{I,DI} = \frac{c_{DI}c_I}{c_{DI}2c^\theta}$$

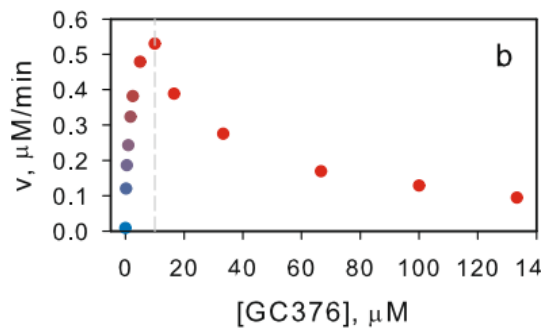
$$K_{S,DS} = \frac{c_{DSCS}}{c_{DS}2c^\theta} \quad K_{I,DS} = \frac{c_{DSC}c_I}{c_{DSIC}c^\theta}$$

$$K_{S,DI} = \frac{c_{DSCS}}{c_{DSI}c^\theta}$$

$$K_d K_{S,D} = K_{S,M} K_{d,MS-M}$$

$$K_d K_{I,D} = K_{I,M} K_{d,MI-M}$$

$$K_{I,D} K_{S,DI} = K_{S,D} K_{I,DS}$$



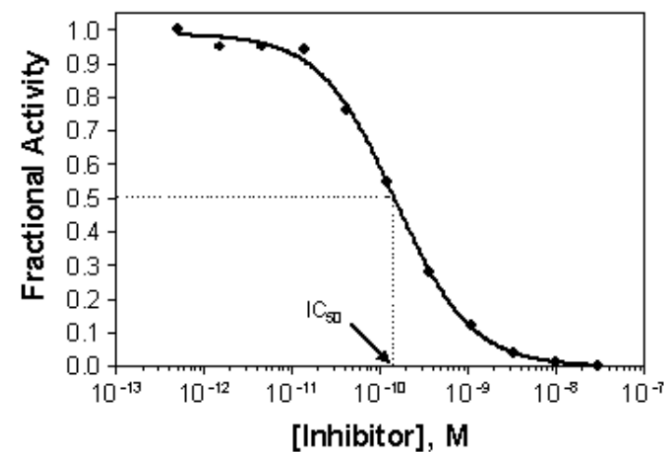
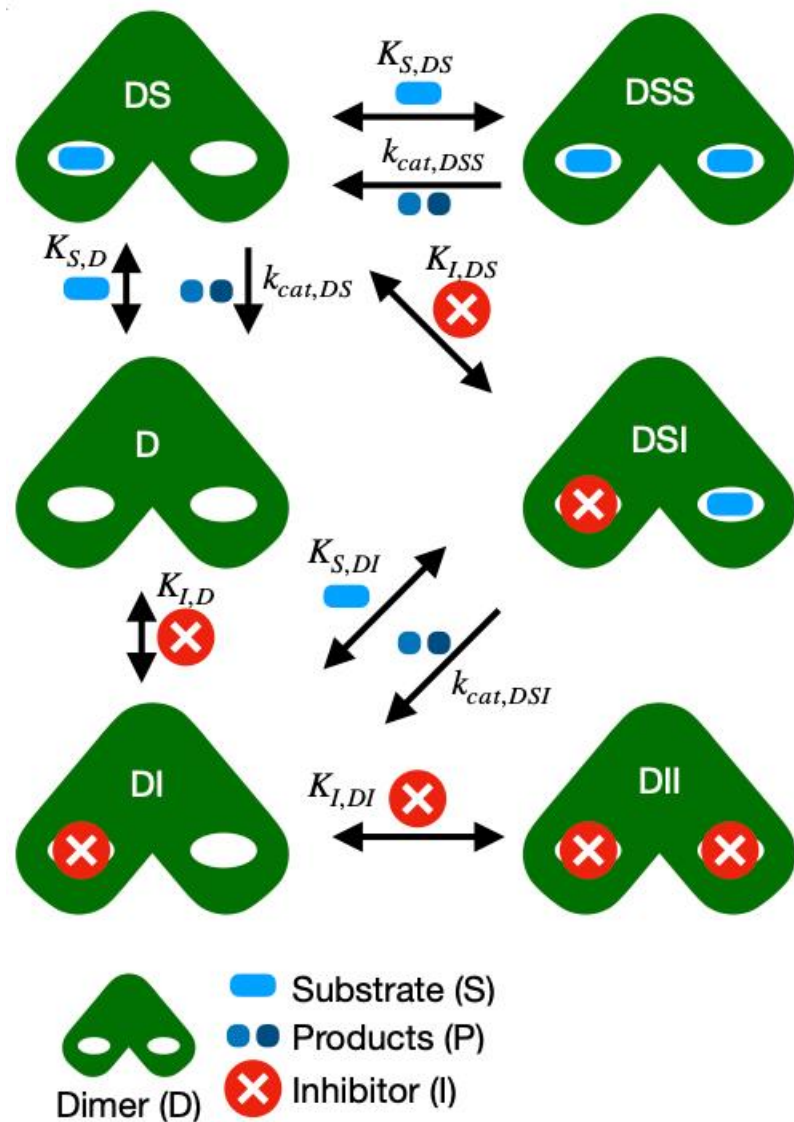
$$c_{Mi} = c_M + c_{MS} + c_{MI} + 2c_D + 2c_{DS} + 2c_{DI} + 2c_{DII} + 2c_{DSS} + 2c_{DSI}$$

$$c_{Si} = c_S + c_{MS} + c_{DS} + 2c_{DSS} + c_{DSI}$$

$$c_{Ii} = c_I + c_{MI} + c_{DI} + 2c_{DII} + c_{DSI}$$

$$v = k_{cat,M} c_{MS} + k_{cat,DS} c_{DS} + k_{cat,DSS} c_{DSS} + k_{cat,DSI} c_{DSI}$$

Prediction based on dimer-only model



It is hypothesized that the dimer-only pIC50 is correlated with cellular pEC50

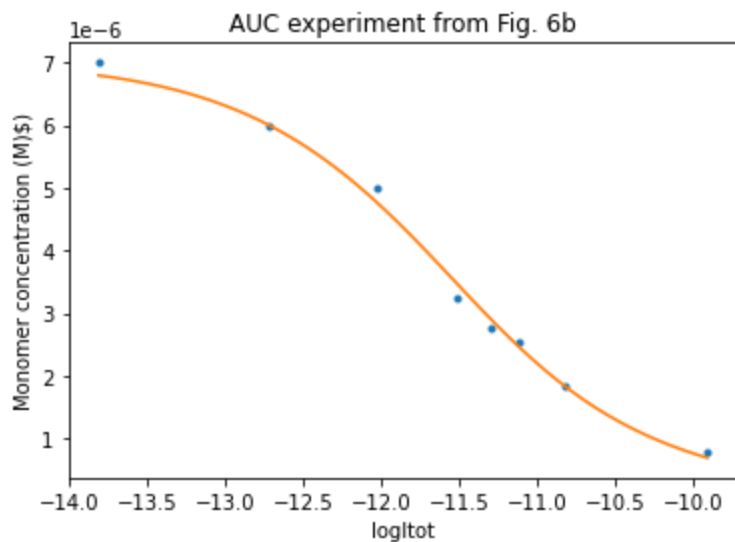
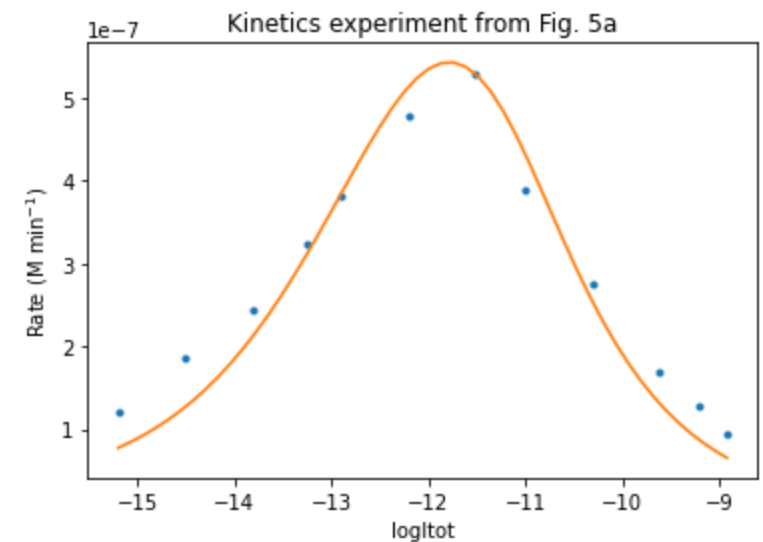
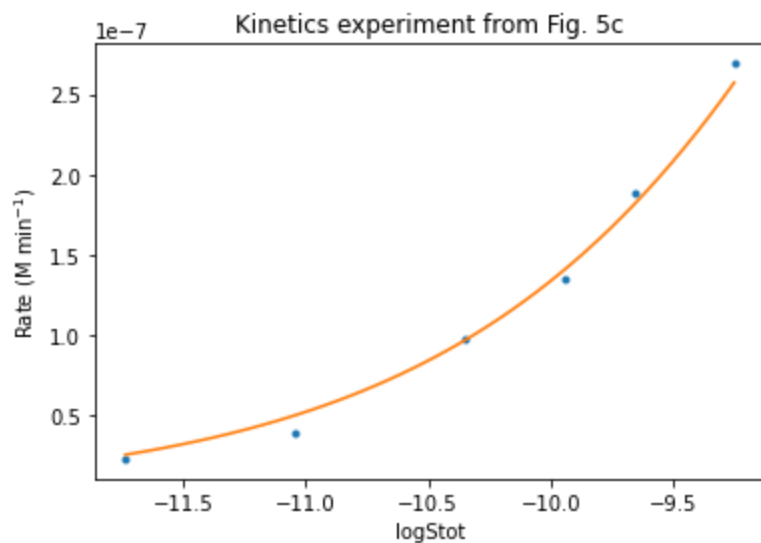
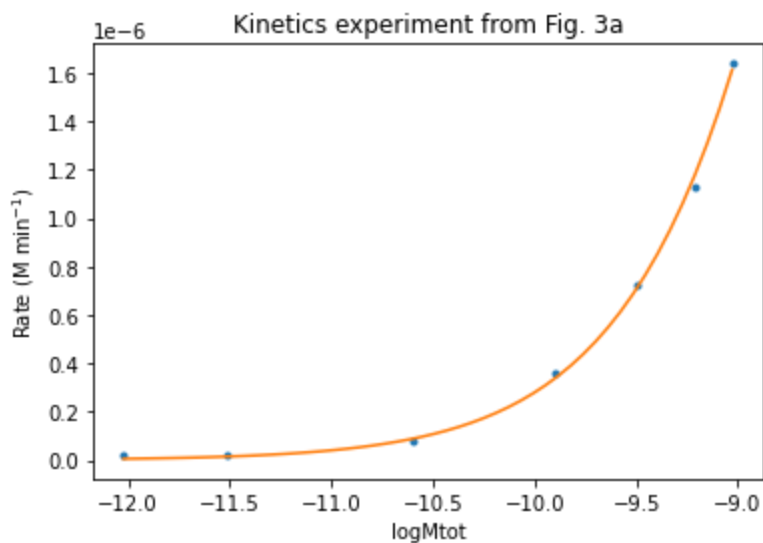
- Aim 1: To construct the Bayesian model for the estimation of kinetic parameters of MPro
- Aim 2: To apply the Bayesian model for the global estimation of kinetic parameters for multiple inhibitors of MPro
- Aim 3: To analyze the correlation between the estimated pIC50 from kinetic model and the cellular pEC50 for multiple inhibitors.

Aim 1: To construct the Bayesian model for the estimation of kinetic parameters

1.1. Constructing Bayesian model for SARS-CoV-2 MPro datasets

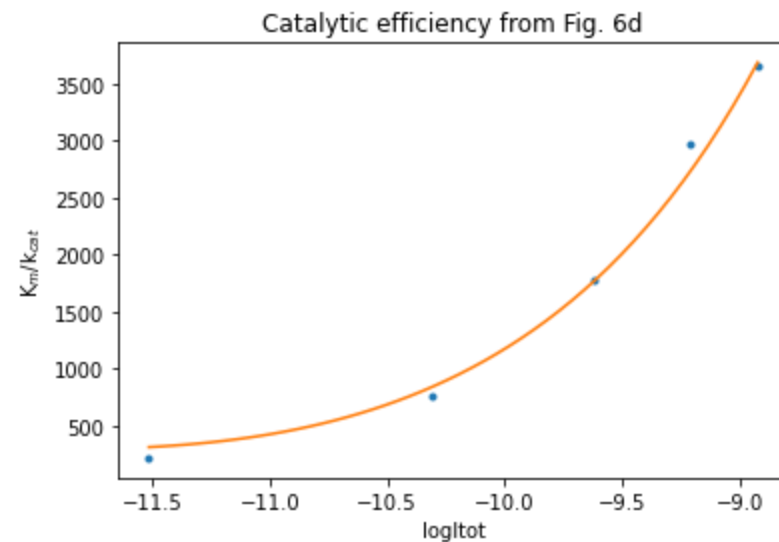
1.2. Simplifying model by adding the constraints on parameters

Nonlinear Regression



logKd	-0.007;	Kd	9.930e-01
logK_S_M	-4.343;	K_S_M	1.300e-02
logK_S_D	-9.313;	K_S_D	9.023e-05
logK_S_DS	-13.898;	K_S_DS	9.211e-07
logK_I_M	-5.121;	K_I_M	5.970e-03
logK_I_D	-17.466;	K_I_D	2.598e-08
logK_I_DI	-18.315;	K_I_DI	1.111e-08
logK_S_DI	-14.234;	K_S_DI	6.580e-07
kcat_MS	0.000e+00		
kcat_DS	0.000e+00		
kcat_DSI	4.094e-01		
kcat_DSS	3.203e-01		

Uncertainty?



Bayesian Regression

Bayes's Theorem:

$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{p(y)} \propto p(\theta)p(y|\theta)$$

- Observations or data $y_i \sim p(y|\theta)$
- Parameter θ

Posterior

Prior

*Likelihood
function*

- + Can incorporate prior knowledge.
- + Better uncertainty quantification.

Data

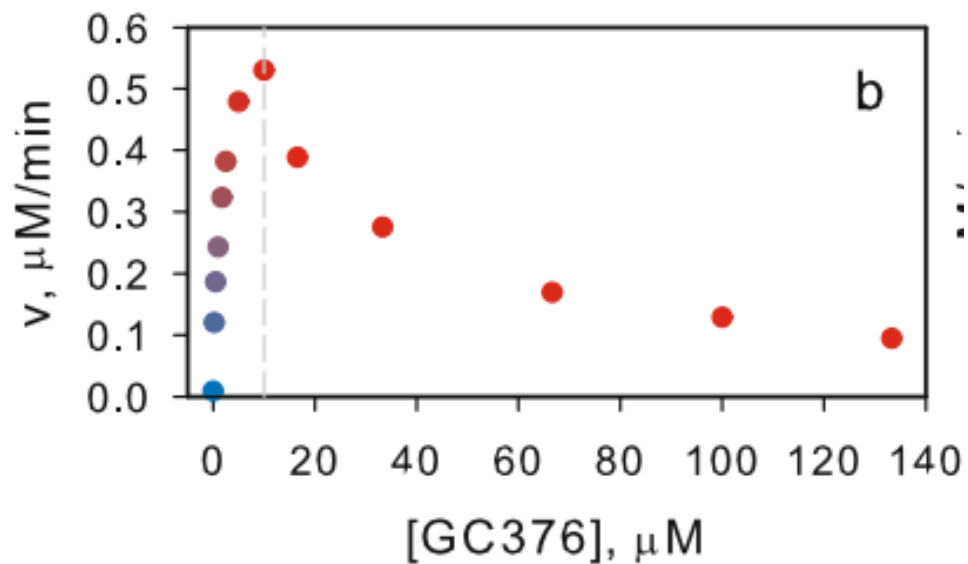


Fig. 5b. Catalytic activity of 10 μM MPro as a function of increasing GC376 concentration.

Modulation of the monomer-dimer equilibrium and catalytic activity of SARS-CoV-2 main protease by a transition-state analog inhibitor

Nashaat T. Nashed¹, Annie Aniana¹, Rodolfo Ghirlando², Sai Chaitanya Chiliveri¹ & John M. Louis¹

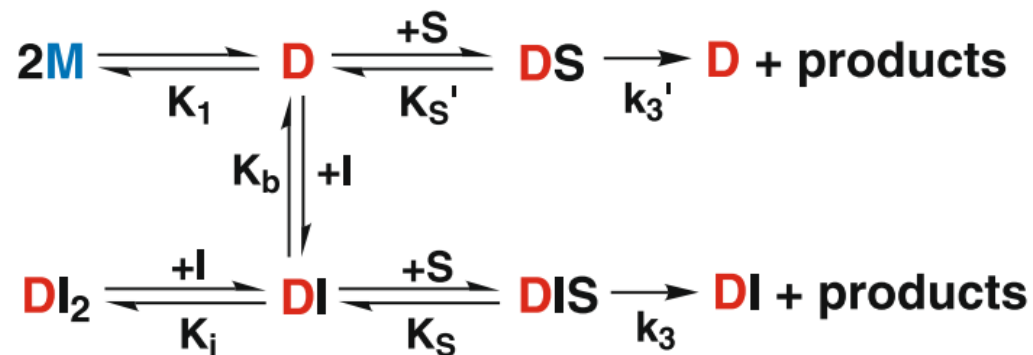
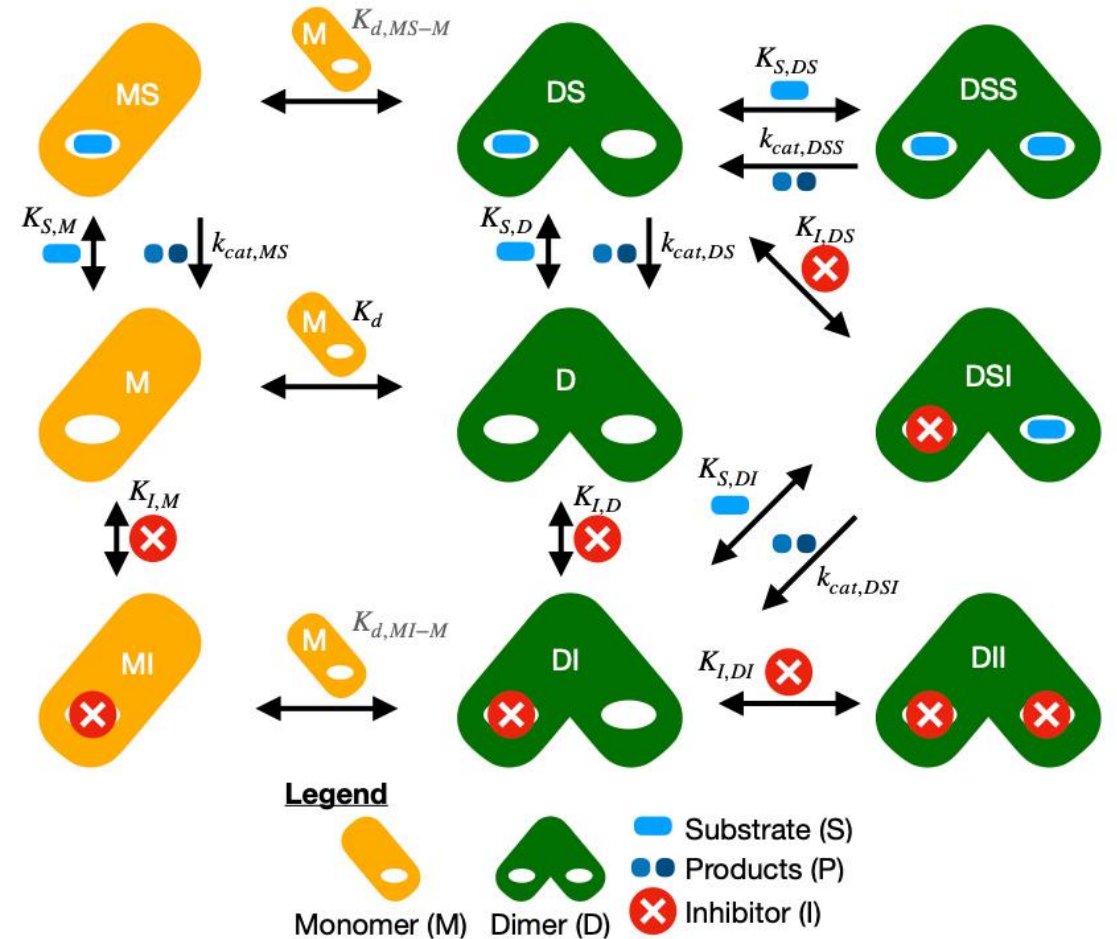


Fig. 8 Mechanism of activation and inhibition of MPro^M by GC376. M, D, S, I, DS, DI, DI₂, DIS denote monomer, dimer, substrate, inhibitor, dimer-substrate complex, dimer-inhibitor complex, dimer bound to 2 inhibitors, dimer bound to 1 inhibitor and 1 substrate, respectively.

Set of parameters

$$\theta \equiv (K_d^i, K_{S,M}, K_{S,D}, K_{S,DS}, K_{I,M}, K_{I,D}, K_{I,DI}, K_{I,DI}, K_{S,DI}, k_{cat,MS}^i, k_{cat,DS}^i, k_{cat,DSS}^i, k_{cat,DSI}^i, k_{cat,DSI}^i, \sigma^j).$$

- K_d and k_{cat} were local parameters, in which i^{th} is the index for MPro^{Mut} and MPro^{WT}.
- The others were the globally shared parameters between the two enzyme variants.
- σ_j is assumed to be constant for the dataset j^{th} .



How can we evaluate the results?

- The convergence of the model (not shown)
- Can model fit all datasets?
- Marginal probability densities
 - 1D - How well do we know each parameters?
 - 2D – How correlated are the parameters?

AIM 1.1: Results

Model fitted
all datasets.

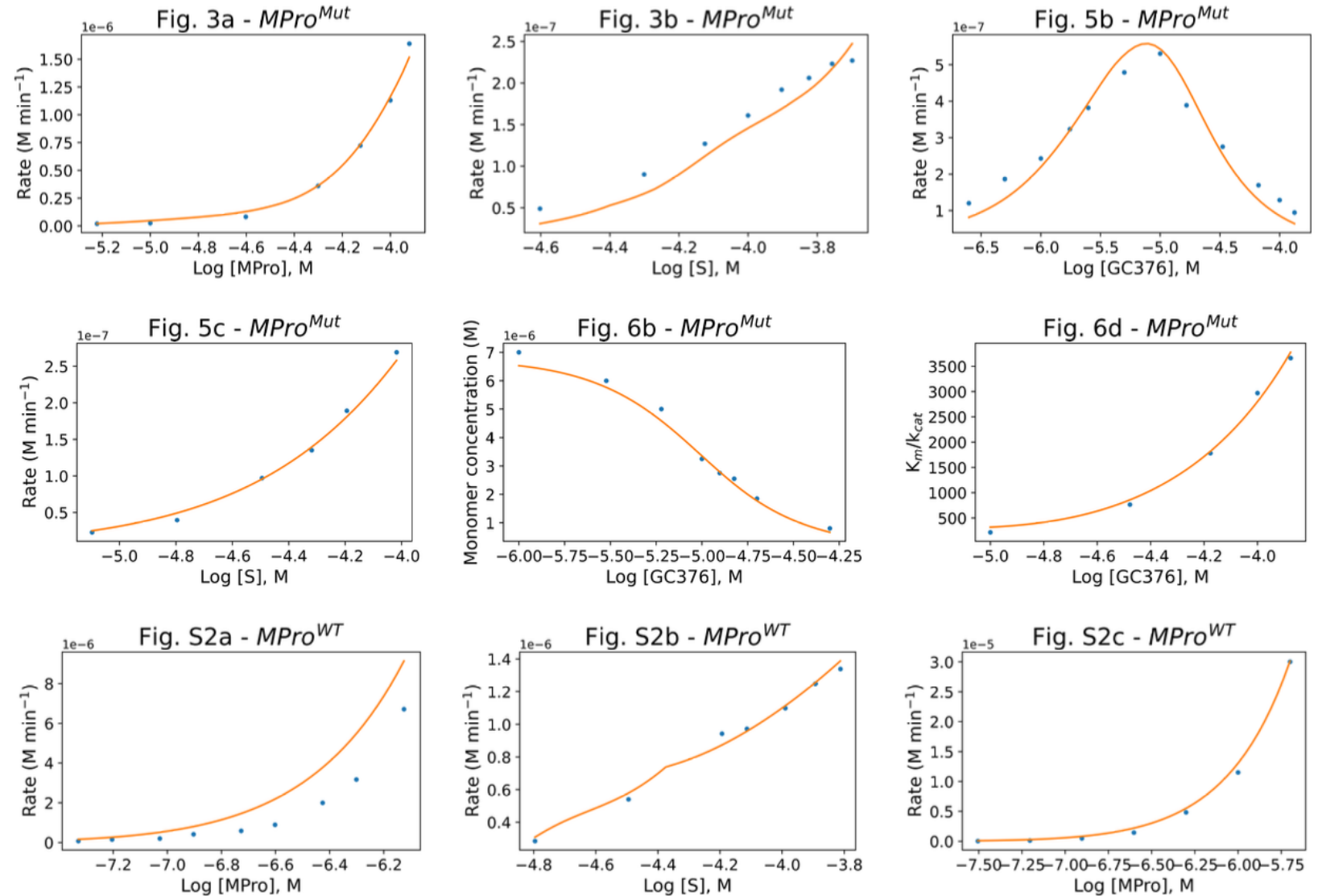
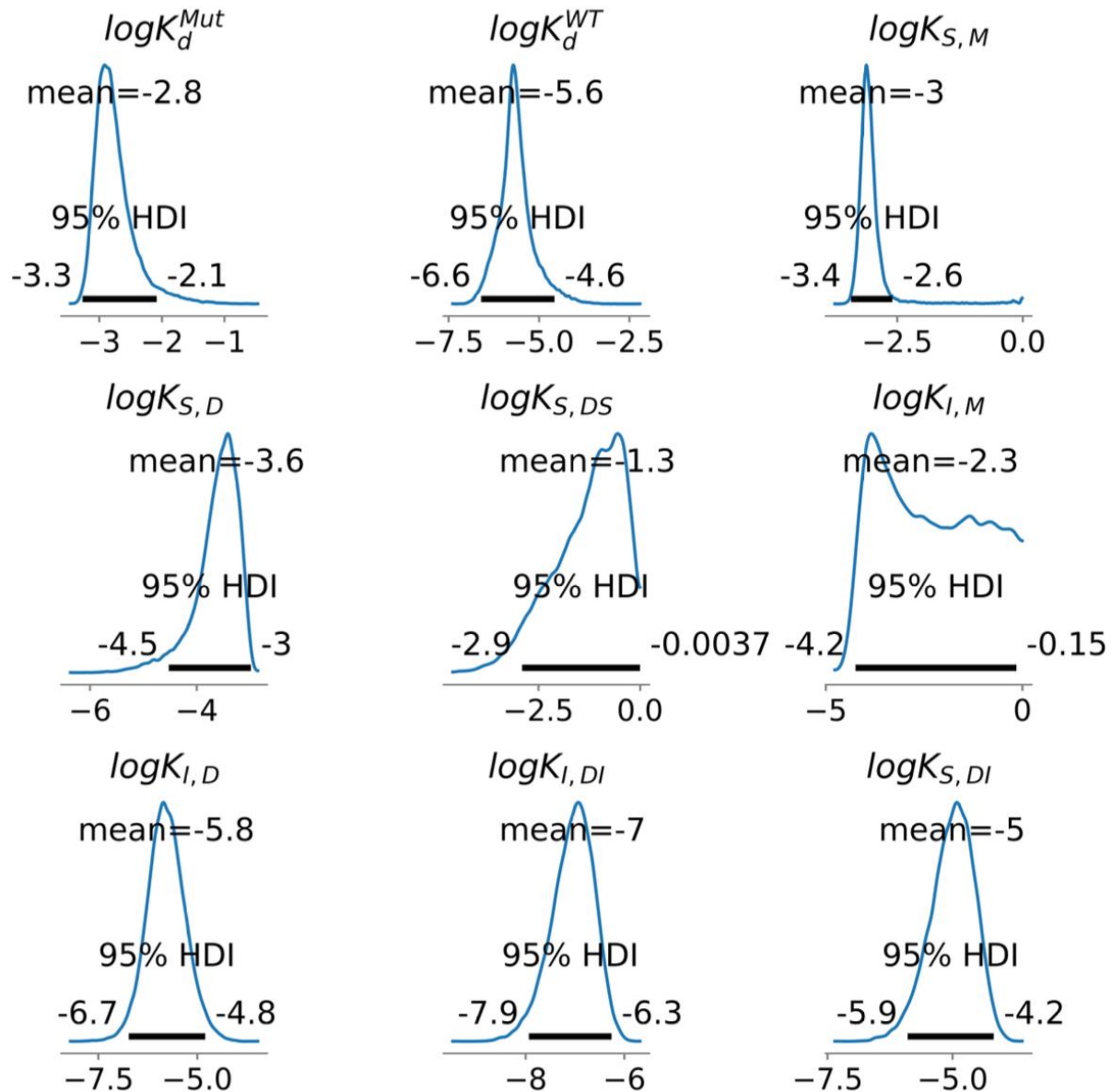


Figure. MAP fitted SARS-CoV-2 datasets

The solid line is the theoretical response $y_n * (\theta_{MAP})$, where θ_{MAP} is the Maximum a Posteriori estimate of the parameters. Dots are the observed response.

(Maximum A Posteriori or MAP: a set of parameter which maximizes the posterior distribution of θ given the data.)

AIM 1.1: Results

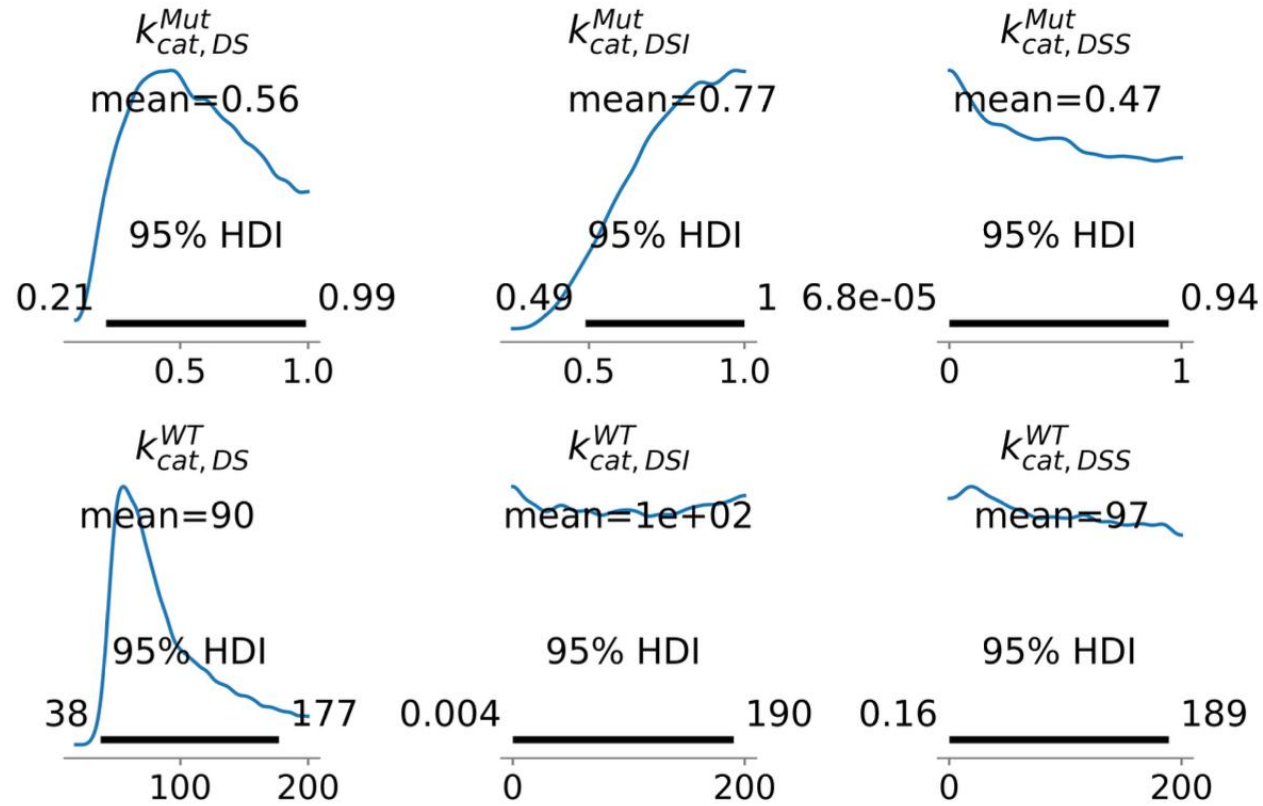


- All $\log K$ parameters are unimodal and have small HDIs, except for $\log K_{I,M}$.
- Large distribution of posterior suggested that this $\log K_{I,M}$ cannot be determined by available datasets.
- K_d^{WT} : [0.04, 18.94] μM
- K_d^{Mut} : [0.43, 7.61] mM

(The **Highest Density Interval** - HDI is the interval which contains the required mass such that all points within the interval have a higher probability density than points outside the interval.)

Figure. Representative 1D marginal distributions of dissociation constants

AIM 1.1: Results



- Large posterior distribution of k_{cat} suggested that these parameters could not be determined by available datasets.

Figure. Representative 1D marginal distributions of rate constants

AIM 1.1: Additional analysis

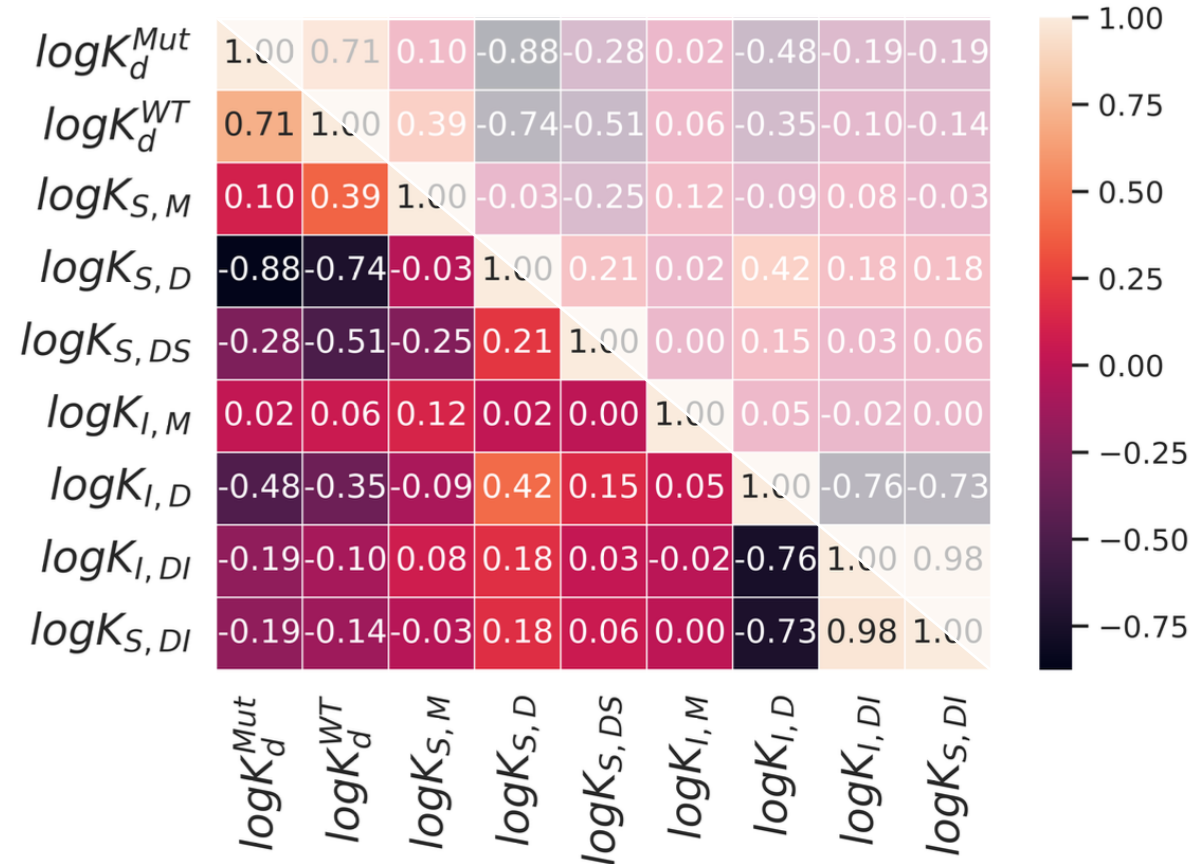
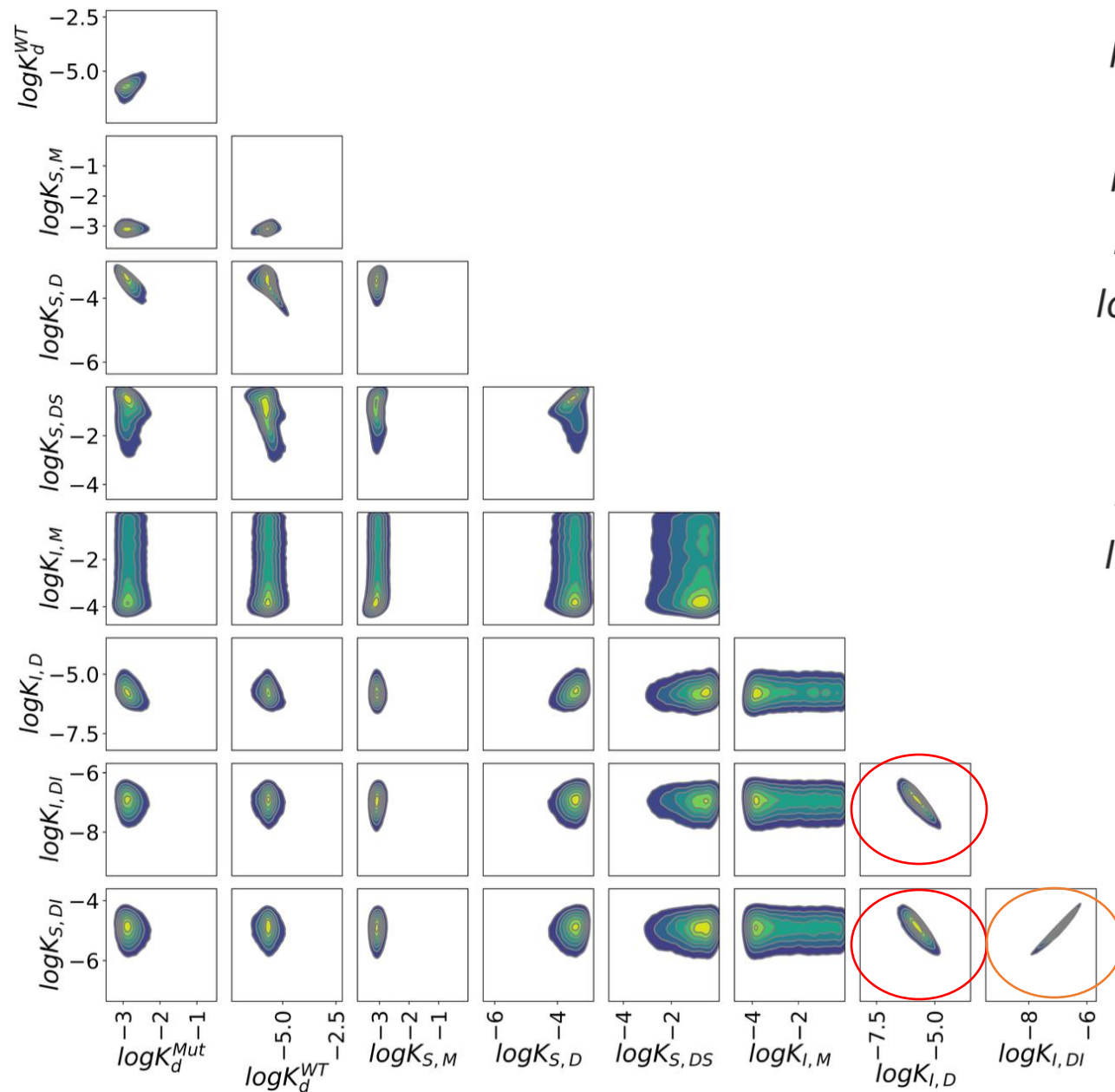


Figure. Heat map of the correlation matrix estimated from the Bayesian posterior of dissociation constants.

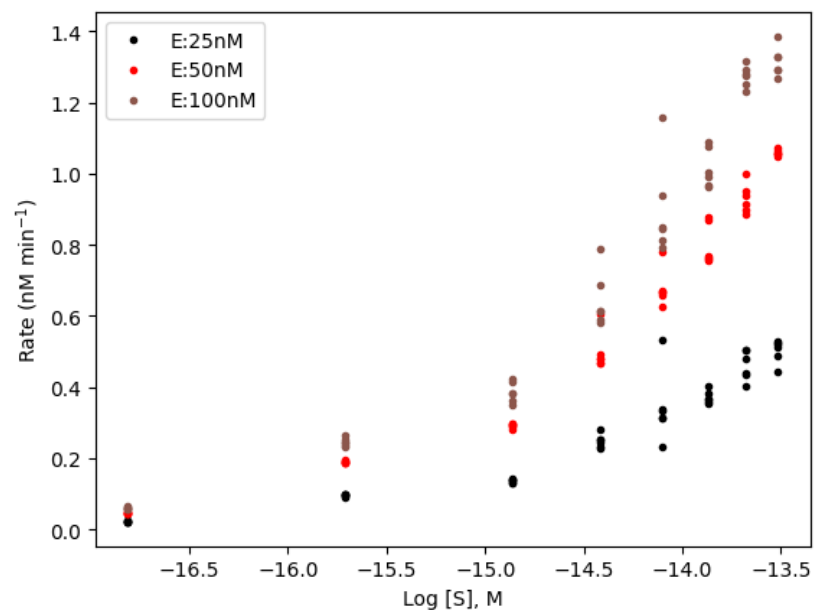
It is hypothesized that estimated pIC50 from kinetic model is correlated with cellular pEC50.

- Aim 1: To construct the Bayesian model for the estimation of kinetic parameters for one set of inhibitor
- Aim 2: To apply the Bayesian model for the global estimation of kinetic parameters for multiple inhibitors
- Aim 3: To analyze the correlation between the estimated pIC50 from kinetic model and the cellular pEC50 for multiple inhibitors.

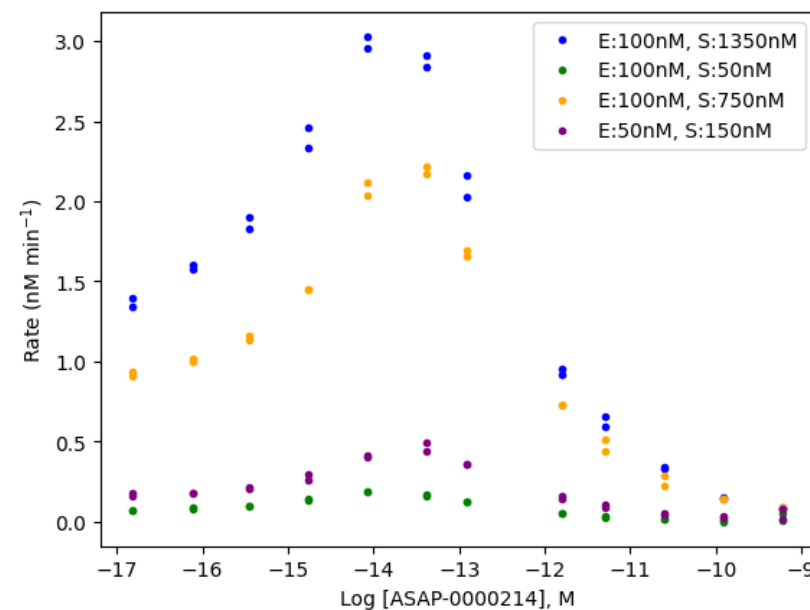
Data from Biochemical assay Core

AI-driven Structure-enabled Antiviral Platform (ASAP)

ASAP uses artificial intelligence and computational chemistry to accelerate structure-based open science antiviral drug discovery and deliver oral antivirals for pandemics with the goal of global, equitable, and affordable access.

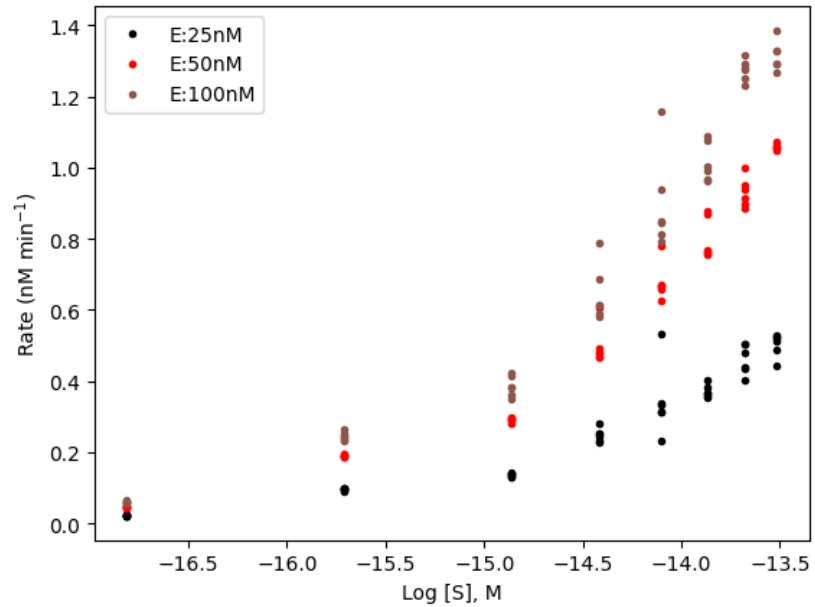


Enzyme – substrate (ES)



Enzyme – substrate – inhibitor (ESI)

AIM 2.1: Methods



Enzyme – substrate (ES)

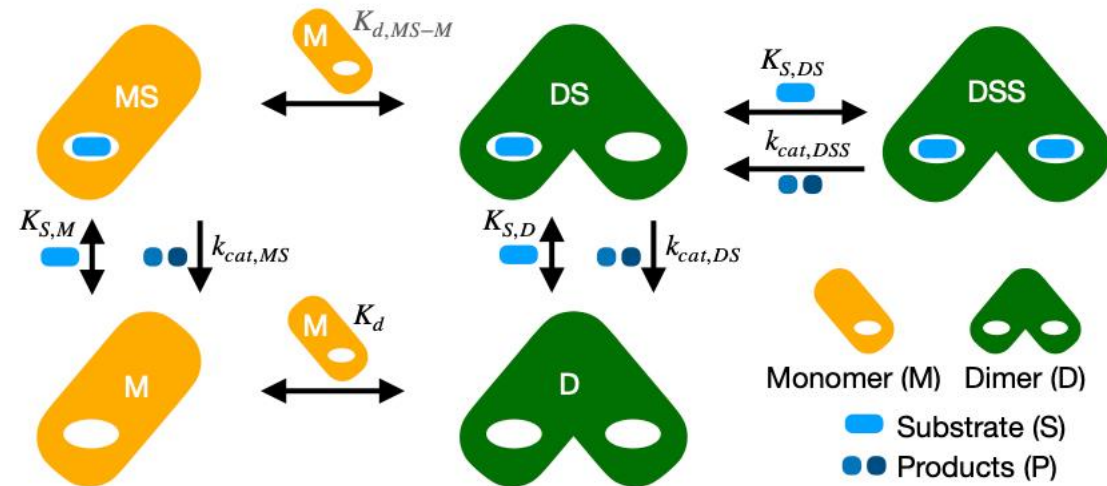
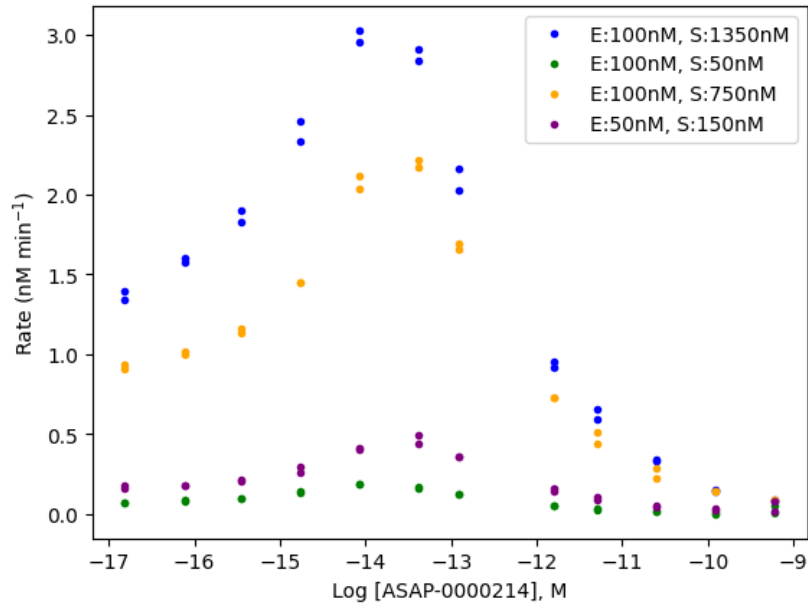


Figure. Enzyme kinetic model with dimerization but without inhibition.

AIM 2.1: Methods



Enzyme – substrate – inhibitor (ESI)

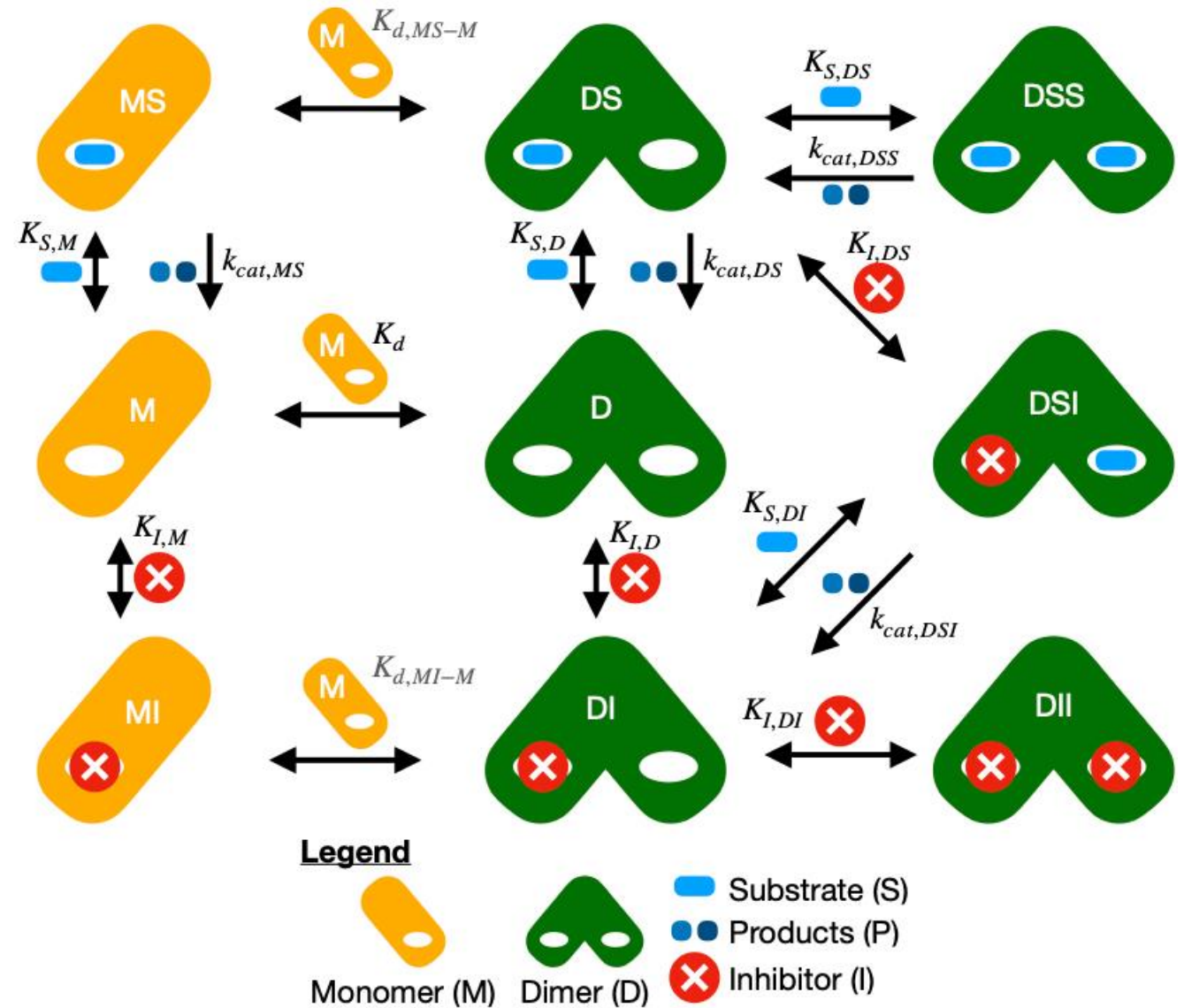
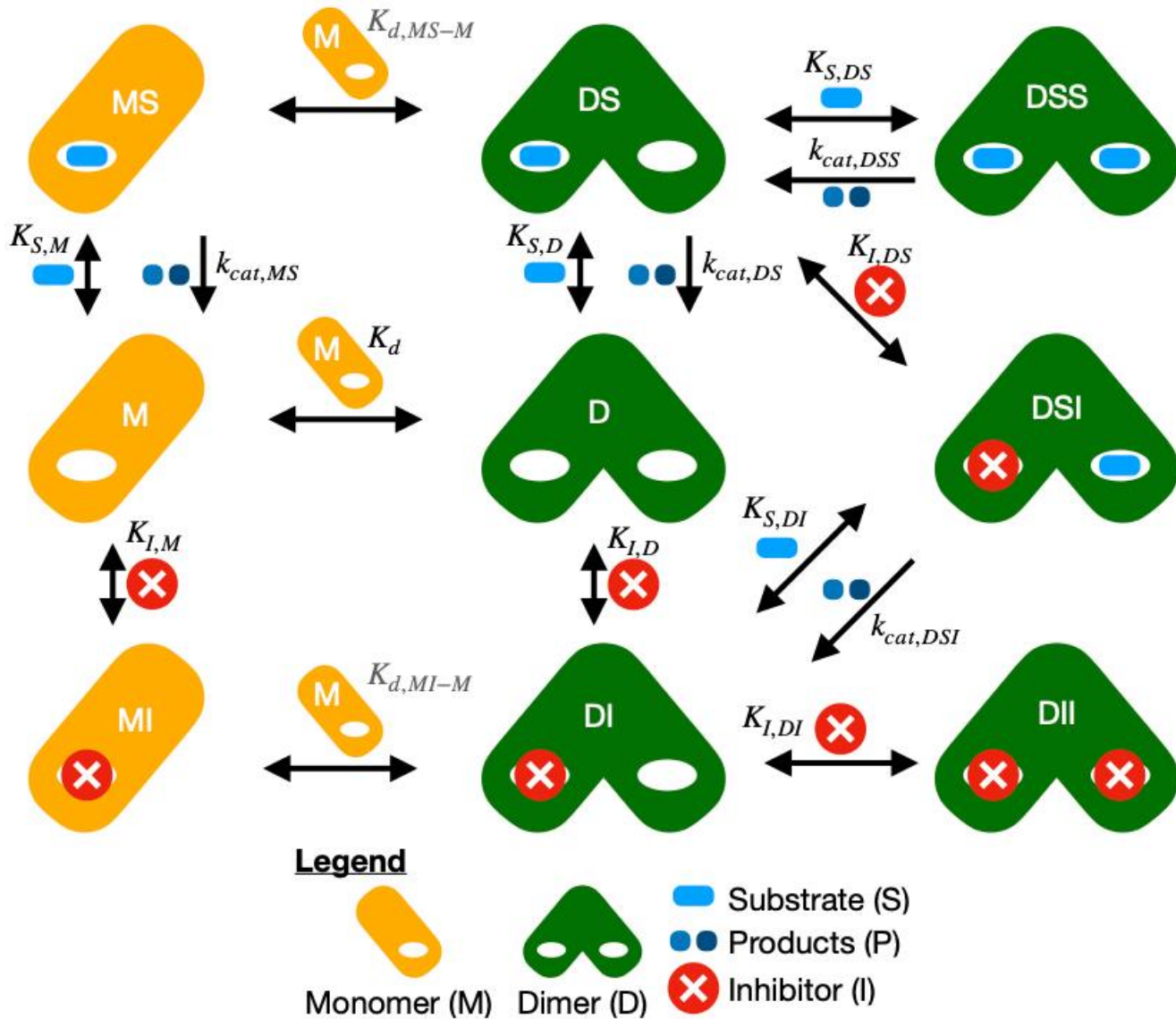


Figure. Full enzyme kinetic model



Dimerization

- $\log K_d$

Binding of substrate (S)

- $\log K_{S,M} + k_{cat,MS}$
- $\log K_{S,D} + k_{cat,DS}$
- $\log K_{S,DS} + k_{cat,DSS}$

Binding of inhibitor (I)

- $\log K_{I,M}$
- $\log K_{I,D}$
- $\log K_{I,DI}$

Binding of S/I

- $\log K_{S,DI} + k_{cat,DSI}$

AIM 2.2: Results

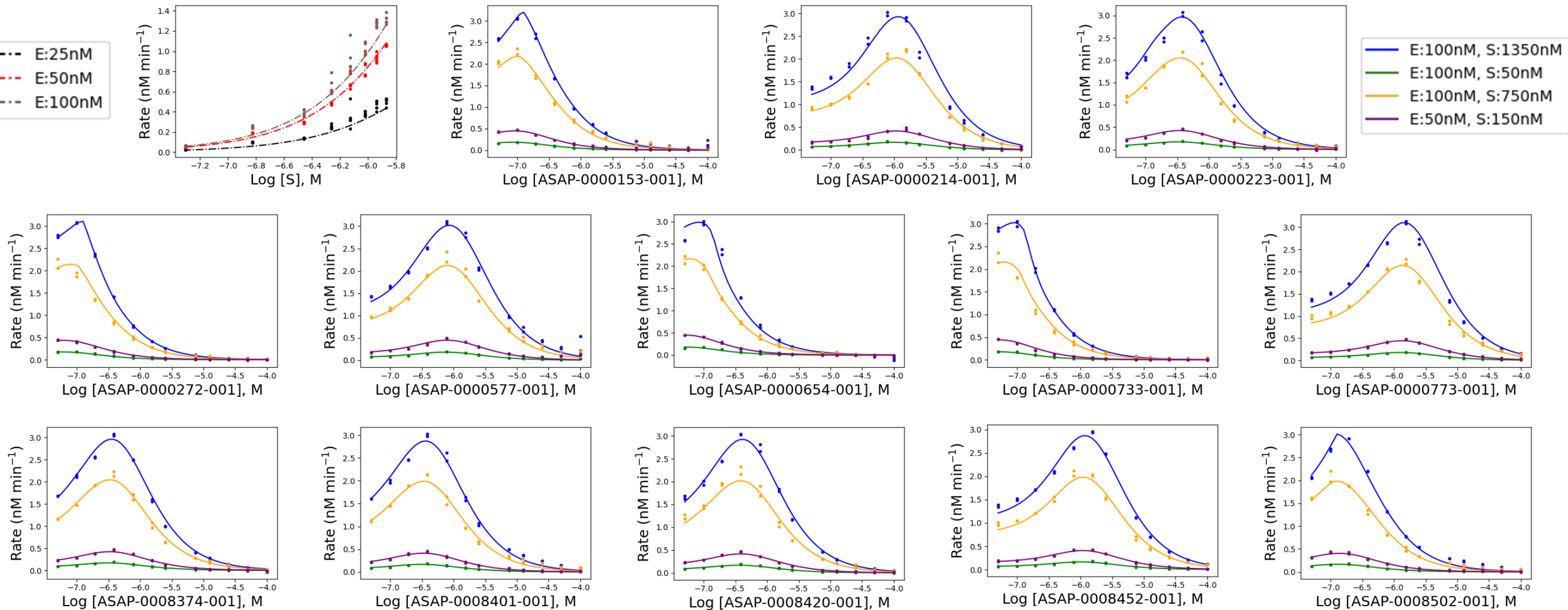


Figure. MAP fitted ES and ESI inhibitor datasets

The lines are the theoretical responses $y_n * (\theta_{MAP})$, where θ_{MAP} is the Maximum a Posteriori estimate of the parameters. Dots are the observed response.

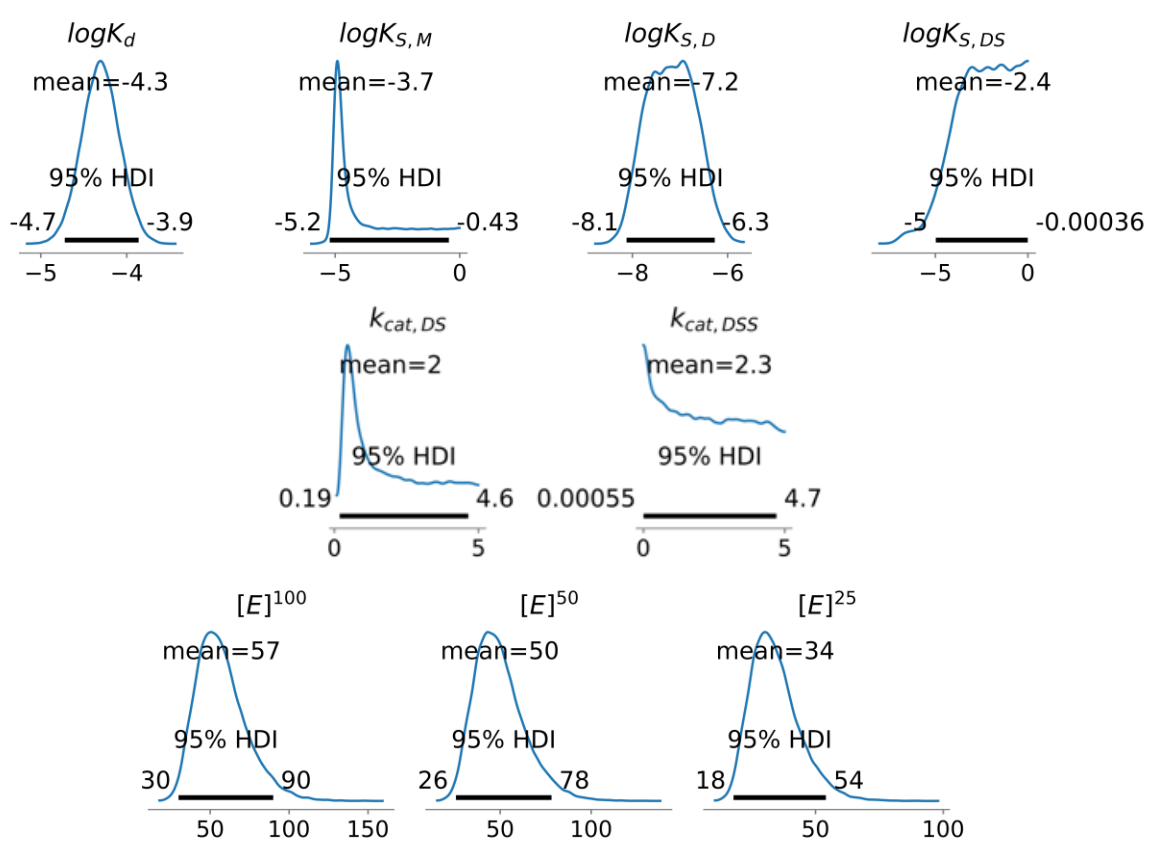


Figure. Representative 1D marginal distributions of parameters from fitting **1 ES and 1 ESI datasets**

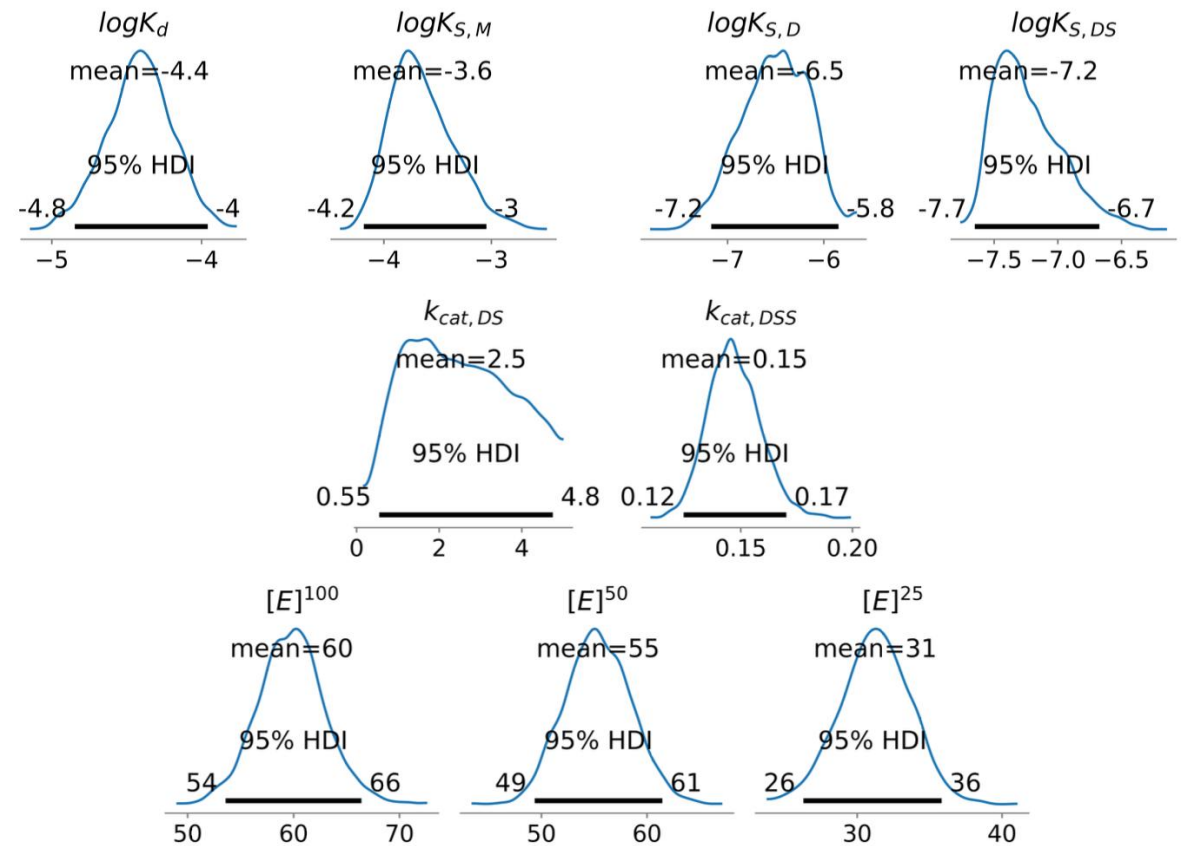
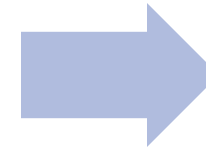
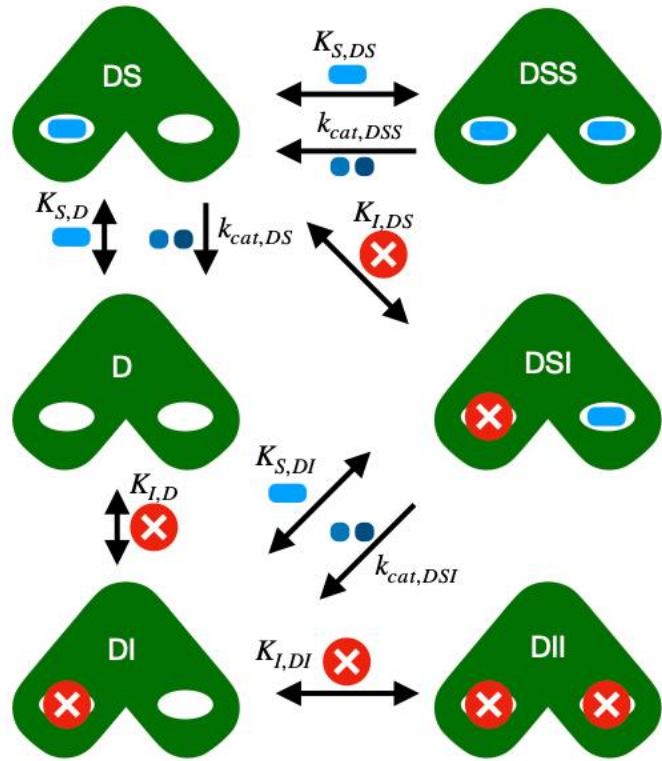


Figure. Representative 1D marginal distributions of shared parameters from **global fitting** (1 ES and 13 ESI datasets)

It is hypothesized that estimated pIC50 from kinetic model is correlated with cellular pEC50.

- Aim 1: To construct the Bayesian model for the estimation of kinetic parameters for one set of inhibitor
- Aim 2: To apply the Bayesian model for the global estimation of kinetic parameters for multiple inhibitors
- Aim 3: To analyze the correlation between the estimated pIC50 from kinetic model and the cellular pEC50 for multiple inhibitors.

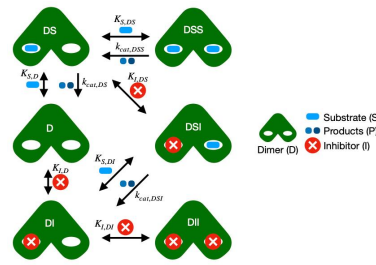
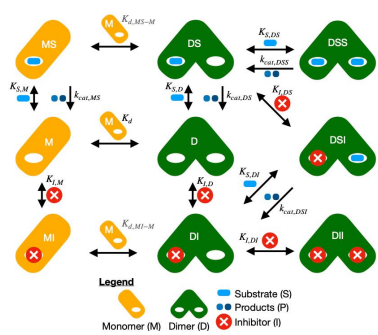
AIM 3: Methods



Estimated
pIC50

Correlation
analysis

Cellular
pEC50



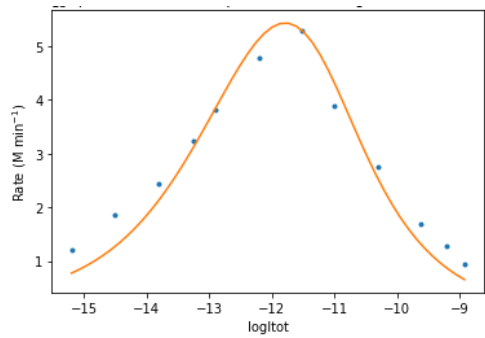
Global fitting
by Bayesian
regression



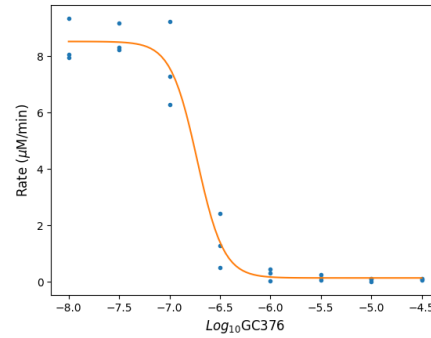
Dimer-only
pIC50 for each
CRC



Correlation
analysis



Fitting 85 CRCs, each curve
specified by a set of enzyme-
inhibitor parameters
(fitting >600 parameters)



$\text{Log}[E] \sim N(\mu=50.0, \sigma=2.5)$
 $\text{Log}[S] \sim N(\mu=1350.0, \sigma=67.5)$
 (unit: nM)

$$y_i = R_{base} + \frac{R_{max} - R_{base}}{1 + 10(\log \text{Concentration}_i - \log I_{C50}) * \text{hill}}$$

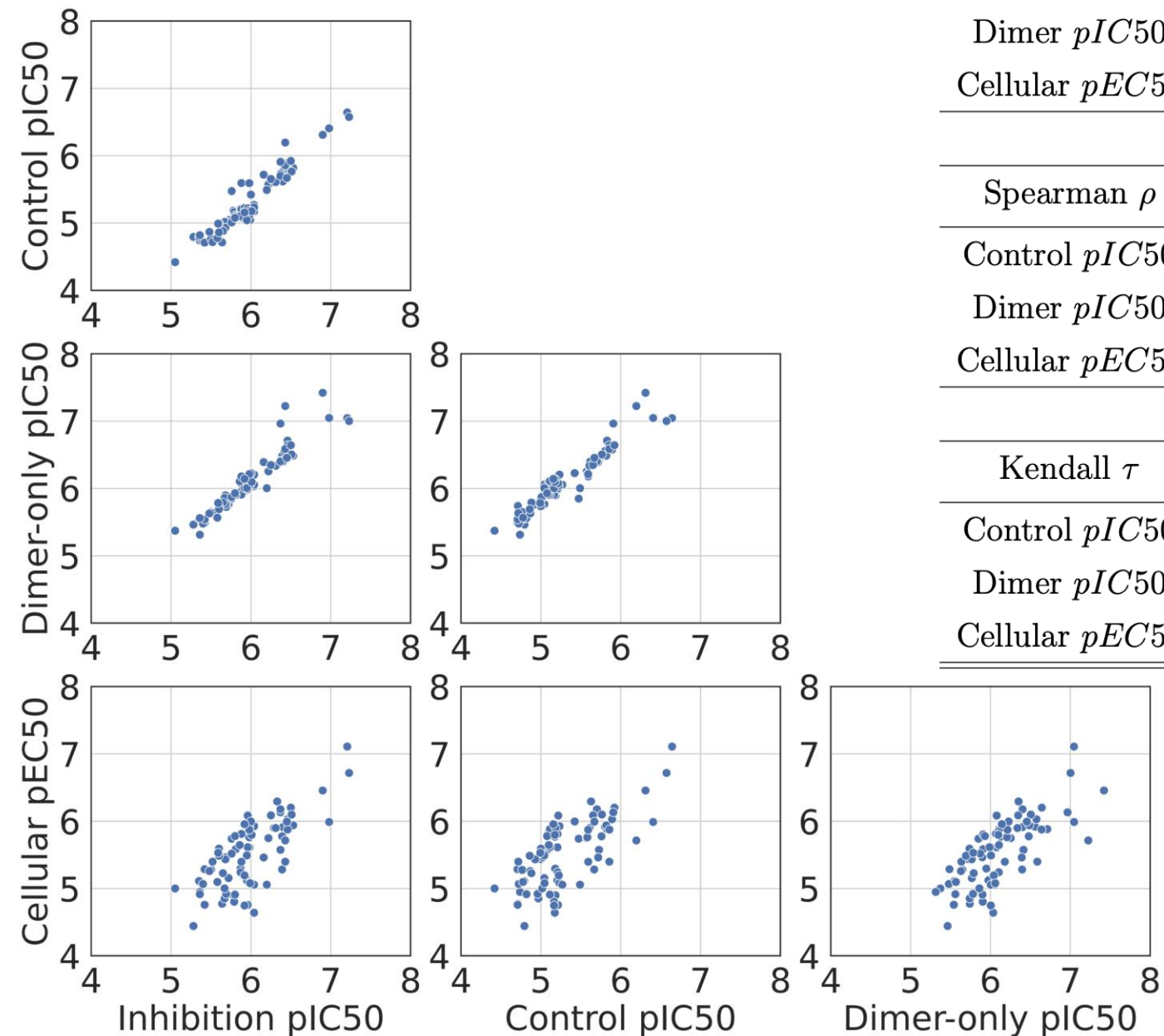
- Pearson R
- Spearman ρ
- Kendall τ
- RMSD

$$\sqrt{\frac{\sum_{i=1}^N (x_i - y_i)^2}{N}}$$

- adjusted RMSD (aRMSD)

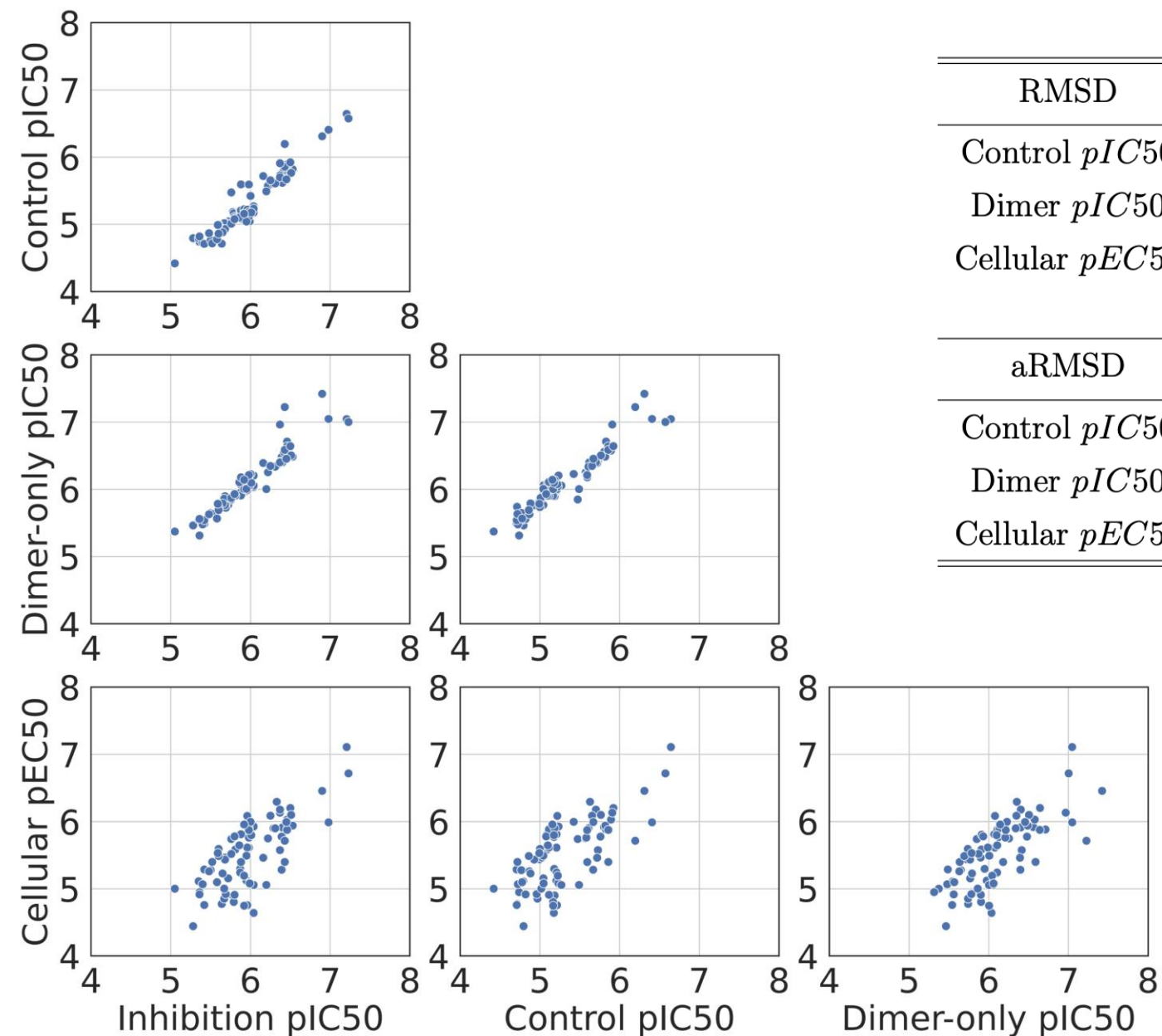
$$\sqrt{\frac{\sum_{i=1}^N (x_i - y_i - (\bar{x} - \bar{y}))^2}{N}}$$

Correlation analysis for pIC50/pEC50



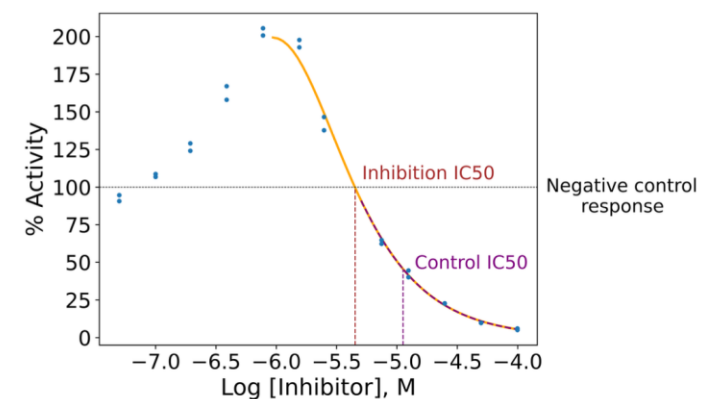
Pearson R	Inhibition $pIC50$	Control $pIC50$	Dimer-only $pIC50$
Control $pIC50$	$0.957 \pm 2.070E-3$		
Dimer $pIC50$	$0.942 \pm 3.941E-2$	$0.947 \pm 3.748E-2$	
Cellular $pEC50$	$0.728 \pm 9.583E-3$	$0.718 \pm 9.682E-3$	$0.726 \pm 1.525E-2$
Spearman ρ	Inhibition $pIC50$	Control $pIC50$	Dimer-only $pIC50$
Control $pIC50$	$0.943 \pm 2.738E-3$		
Dimer $pIC50$	$0.961 \pm 9.801E-3$	$0.951 \pm 8.484E-3$	
Cellular $pEC50$	$0.680 \pm 9.347E-3$	$0.665 \pm 8.970E-3$	$0.746 \pm 2.895E-3$
Kendall τ	Inhibition $pIC50$	Control $pIC50$	Dimer-only $pIC50$
Control $pIC50$	$0.808 \pm 4.384E-3$		
Dimer $pIC50$	$0.852 \pm 7.552E-3$	$0.829 \pm 6.225E-3$	
Cellular $pEC50$	$0.499 \pm 8.156E-3$	$0.488 \pm 7.820E-3$	$0.546 \pm 3.244E-3$

Correlation analysis for pIC50/pEC50

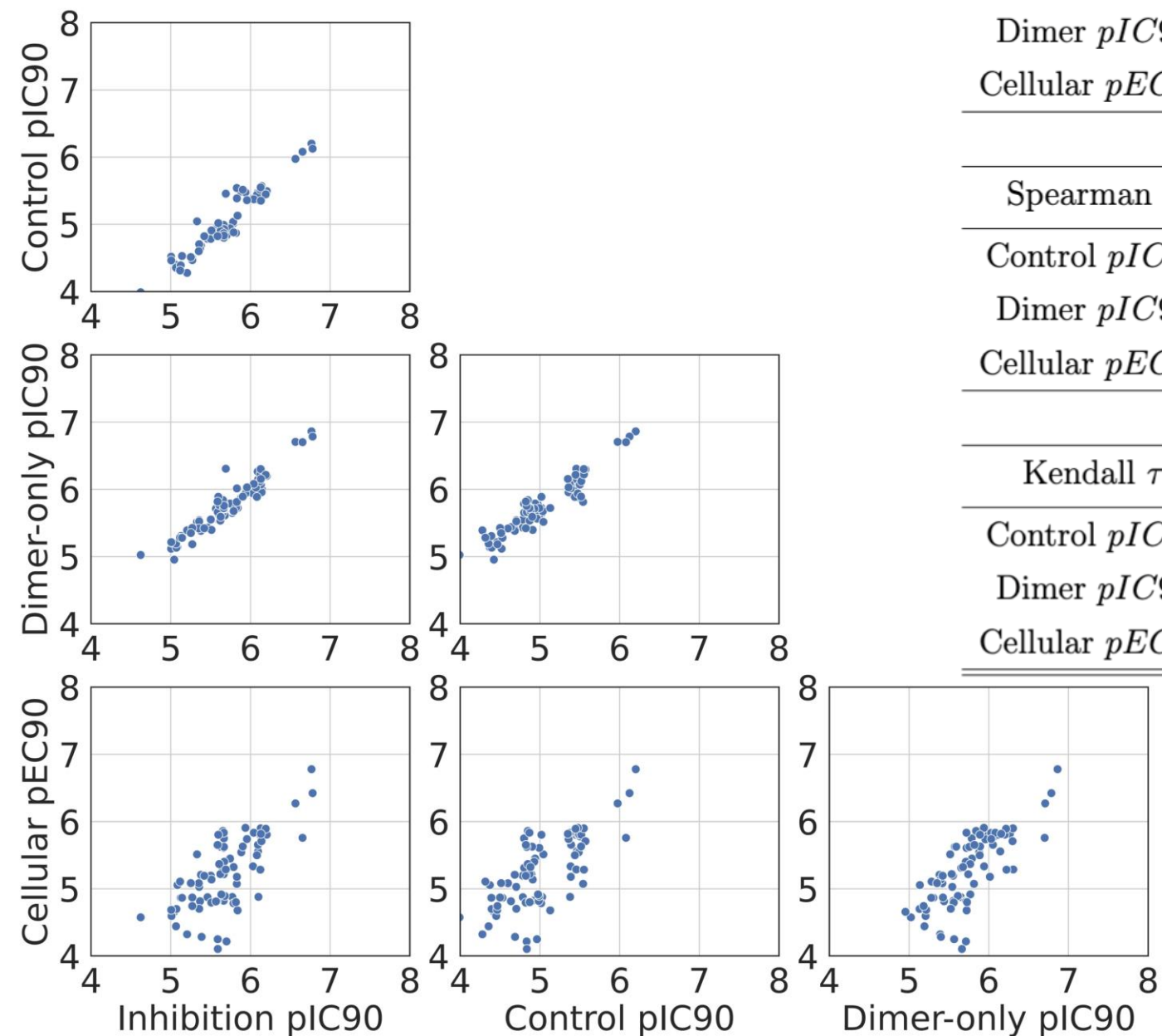


RMSD	Inhibition pIC_{50}	Control pIC_{50}	Dimer-only pIC_{50}
Control pIC_{50}	$0.686 \pm 2.222E-3$		
Dimer pIC_{50}	$0.186 \pm 3.482E-2$	$0.804 \pm 2.350E-2$	
Cellular pEC_{50}	$0.574 \pm 6.836E-3$	$0.417 \pm 3.670E-3$	$0.668 \pm 2.083E-2$

aRMSD	Inhibition pIC_{50}	Control pIC_{50}	Dimer-only pIC_{50}
Control pIC_{50}	$0.133 \pm 2.462E-3$		
Dimer pIC_{50}	$0.136 \pm 4.051E-2$	$0.161 \pm 3.648E-2$	
Cellular pEC_{50}	$0.371 \pm 3.715E-3$	$0.379 \pm 3.761E-3$	$0.339 \pm 1.552E-2$



Correlation analysis for pIC90/pEC90

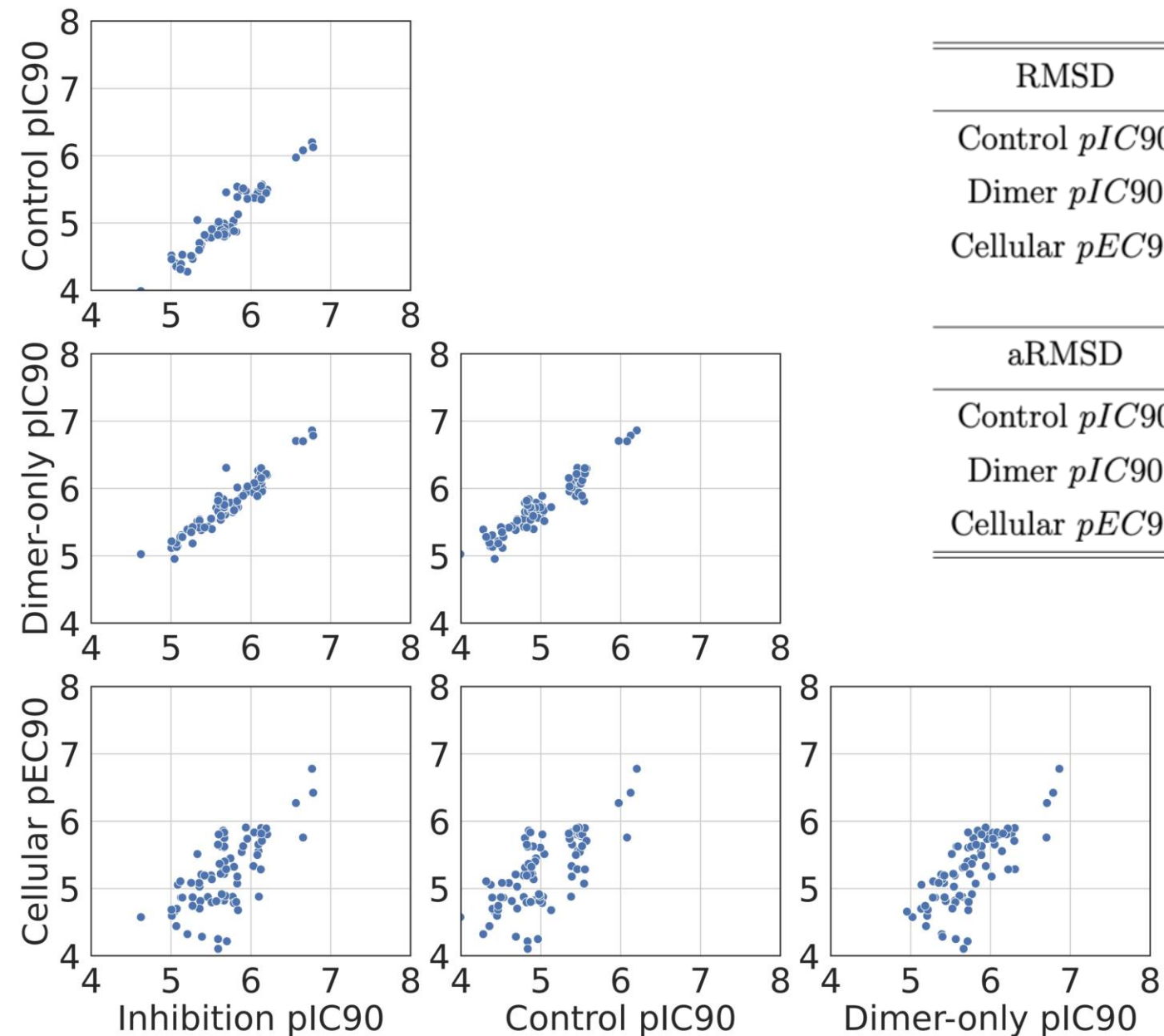


Pearson R	Inhibition $pIC90$	Control $pIC90$	Dimer-only $pIC90$
Control $pIC90$	$0.955 \pm 2.083E-3$		
Dimer $pIC90$	$0.942 \pm 4.352E-2$	$0.933 \pm 4.102E-2$	
Cellular $pEC90$	$0.682 \pm 1.171E-2$	$0.671 \pm 1.139E-2$	$0.724 \pm 1.657E-2$

Spearman ρ	Inhibition $pIC90$	Control $pIC90$	Dimer-only $pIC90$
Control $pIC90$	$0.920 \pm 3.332E-3$		
Dimer $pIC90$	$0.933 \pm 9.812E-3$	$0.900 \pm 7.380E-3$	
Cellular $pEC90$	$0.686 \pm 9.496E-3$	$0.648 \pm 1.013E-2$	$0.762 \pm 2.016E-3$

Kendall τ	Inhibition $pIC90$	Control $pIC90$	Dimer-only $pIC90$
Control $pIC90$	$0.759 \pm 5.367E-3$		
Dimer $pIC90$	$0.793 \pm 6.840E-3$	$0.725 \pm 5.963E-3$	
Cellular $pEC90$	$0.500 \pm 8.400E-3$	$0.465 \pm 8.776E-3$	$0.561 \pm 2.015E-3$

Correlation analysis for pIC90/pEC90



RMSD	Inhibition $pIC90$	Control $pIC90$	Dimer-only $pIC90$
Control $pIC90$	$0.686 \pm 2.222E-3$		
Dimer $pIC90$	$0.148 \pm 3.973E-2$	$0.744 \pm 2.108E-2$	
Cellular $pEC90$	$0.633 \pm 1.151E-2$	$0.482 \pm 6.063E-3$	$0.654 \pm 1.740E-2$

aRMSD	Inhibition $pIC90$	Control $pIC90$	Dimer-only $pIC90$
Control $pIC90$	$0.133 \pm 2.462E-3$		
Dimer $pIC90$	$0.136 \pm 4.051E-2$	$0.161 \pm 3.648E-2$	
Cellular $pEC90$	$0.427 \pm 9.135E-3$	$0.436 \pm 8.550E-3$	$0.405 \pm 1.068E-2$

Pearson R	Inhibition $pIC50$	Control $pIC50$	Dimer-only $pIC50$
Control $pIC50$	$0.957 \pm 2.070E-3$		
Dimer $pIC50$	$0.942 \pm 3.941E-2$	$0.947 \pm 3.748E-2$	
Cellular $pEC50$	$0.728 \pm 9.583E-3$	$0.718 \pm 9.682E-3$	$0.726 \pm 1.525E-2$

Spearman ρ	Inhibition $pIC50$	Control $pIC50$	Dimer-only $pIC50$
Control $pIC50$	$0.943 \pm 2.738E-3$		
Dimer $pIC50$	$0.961 \pm 9.801E-3$	$0.951 \pm 8.484E-3$	
Cellular $pEC50$	$0.680 \pm 9.347E-3$	$0.665 \pm 8.970E-3$	$0.746 \pm 2.895E-3$

Kendall τ	Inhibition $pIC50$	Control $pIC50$	Dimer-only $pIC50$
Control $pIC50$	$0.808 \pm 4.384E-3$		
Dimer $pIC50$	$0.852 \pm 7.552E-3$	$0.829 \pm 6.225E-3$	
Cellular $pEC50$	$0.499 \pm 8.156E-3$	$0.488 \pm 7.820E-3$	$0.546 \pm 3.244E-3$

RMSD	Inhibition $pIC50$	Control $pIC50$	Dimer-only $pIC50$
Control $pIC50$	$0.686 \pm 2.222E-3$		
Dimer $pIC50$	$0.186 \pm 3.482E-2$	$0.804 \pm 2.350E-2$	
Cellular $pEC50$	$0.574 \pm 6.836E-3$	$0.417 \pm 3.670E-3$	$0.668 \pm 2.083E-2$

aRMSD	Inhibition $pIC50$	Control $pIC50$	Dimer-only $pIC50$
Control $pIC50$	$0.133 \pm 2.462E-3$		
Dimer $pIC50$	$0.136 \pm 4.051E-2$	$0.161 \pm 3.648E-2$	
Cellular $pEC50$	$0.371 \pm 3.715E-3$	$0.379 \pm 3.761E-3$	$0.339 \pm 1.552E-2$

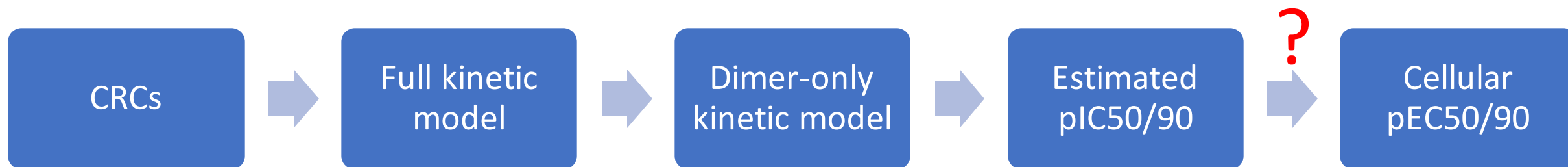
Pearson R	Inhibition $pIC90$	Control $pIC90$	Dimer-only $pIC90$
Control $pIC90$	$0.955 \pm 2.083E-3$		
Dimer $pIC90$	$0.942 \pm 4.352E-2$	$0.933 \pm 4.102E-2$	
Cellular $pEC90$	$0.682 \pm 1.171E-2$	$0.671 \pm 1.139E-2$	$0.724 \pm 1.657E-2$

Spearman ρ	Inhibition $pIC90$	Control $pIC90$	Dimer-only $pIC90$
Control $pIC90$	$0.920 \pm 3.332E-3$		
Dimer $pIC90$	$0.933 \pm 9.812E-3$	$0.900 \pm 7.380E-3$	
Cellular $pEC90$	$0.686 \pm 9.496E-3$	$0.648 \pm 1.013E-2$	$0.762 \pm 2.016E-3$

Kendall τ	Inhibition $pIC90$	Control $pIC90$	Dimer-only $pIC90$
Control $pIC90$	$0.759 \pm 5.367E-3$		
Dimer $pIC90$	$0.793 \pm 6.840E-3$	$0.725 \pm 5.963E-3$	
Cellular $pEC90$	$0.500 \pm 8.400E-3$	$0.465 \pm 8.776E-3$	$0.561 \pm 2.015E-3$

RMSD	Inhibition $pIC90$	Control $pIC90$	Dimer-only $pIC90$
Control $pIC90$	$0.686 \pm 2.222E-3$		
Dimer $pIC90$	$0.148 \pm 3.973E-2$	$0.744 \pm 2.108E-2$	
Cellular $pEC90$	$0.633 \pm 1.151E-2$	$0.482 \pm 6.063E-3$	$0.654 \pm 1.740E-2$

aRMSD	Inhibition $pIC90$	Control $pIC90$	Dimer-only $pIC90$
Control $pIC90$	$0.133 \pm 2.462E-3$		
Dimer $pIC90$	$0.136 \pm 4.051E-2$	$0.161 \pm 3.648E-2$	
Cellular $pEC90$	$0.427 \pm 9.135E-3$	$0.436 \pm 8.550E-3$	$0.405 \pm 1.068E-2$



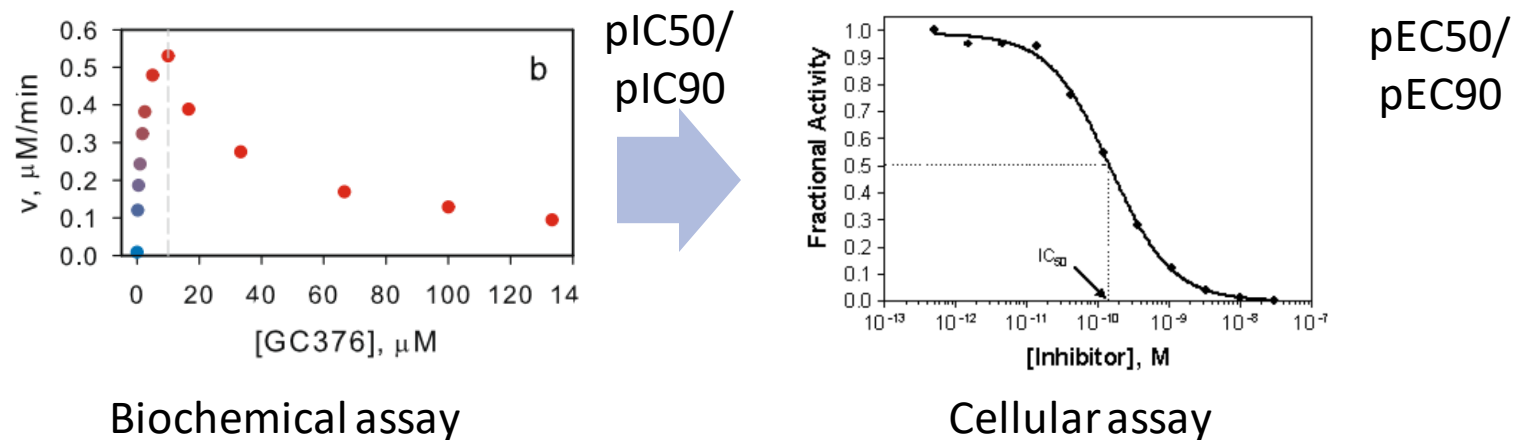
Dimer-only pIC50/90 estimated from biphasic curve are highly correlated with cellular pEC50/90

Summary

- We successfully constructed the kinetic model to estimate kinetic parameters of MPro from SARS-CoV-2 and MERS-CoV.
- Dimer-only pIC50/90 estimated from biphasic curve are highly correlated with cellular pEC50/90
- The accuracy of predicting cellular pEC90 using dimer-only pIC90 was higher than that of predicting pEC50 using dimer-only pIC50.

Contribution




- Given the biphasic concentration-response curves, estimated pIC50/pIC90 from dimer-only model can be used for compound screening and aid the drug design.



- The model can be adjusted and applied to any system with complex mechanisms involving dimeric enzymes and numerous binding events.

Acknowledgements

Modulation of the monomer-dimer equilibrium and catalytic activity of SARS-CoV-2 main protease by a transition-state analog inhibitor

Nashaat T. Nashed¹, Annie Aniana¹, Rodolfo Ghirlando ², Sai Chaitanya Chiliveri ¹ & John M. Louis ¹✉

AI-driven Structure-enabled Antiviral Platform (ASAP)

ASAP uses artificial intelligence and computational chemistry to accelerate structure-based open science antiviral drug discovery and deliver oral antivirals for pandemics with the goal of global, equitable, and affordable access.

**Thank you for your
attention!!!!**

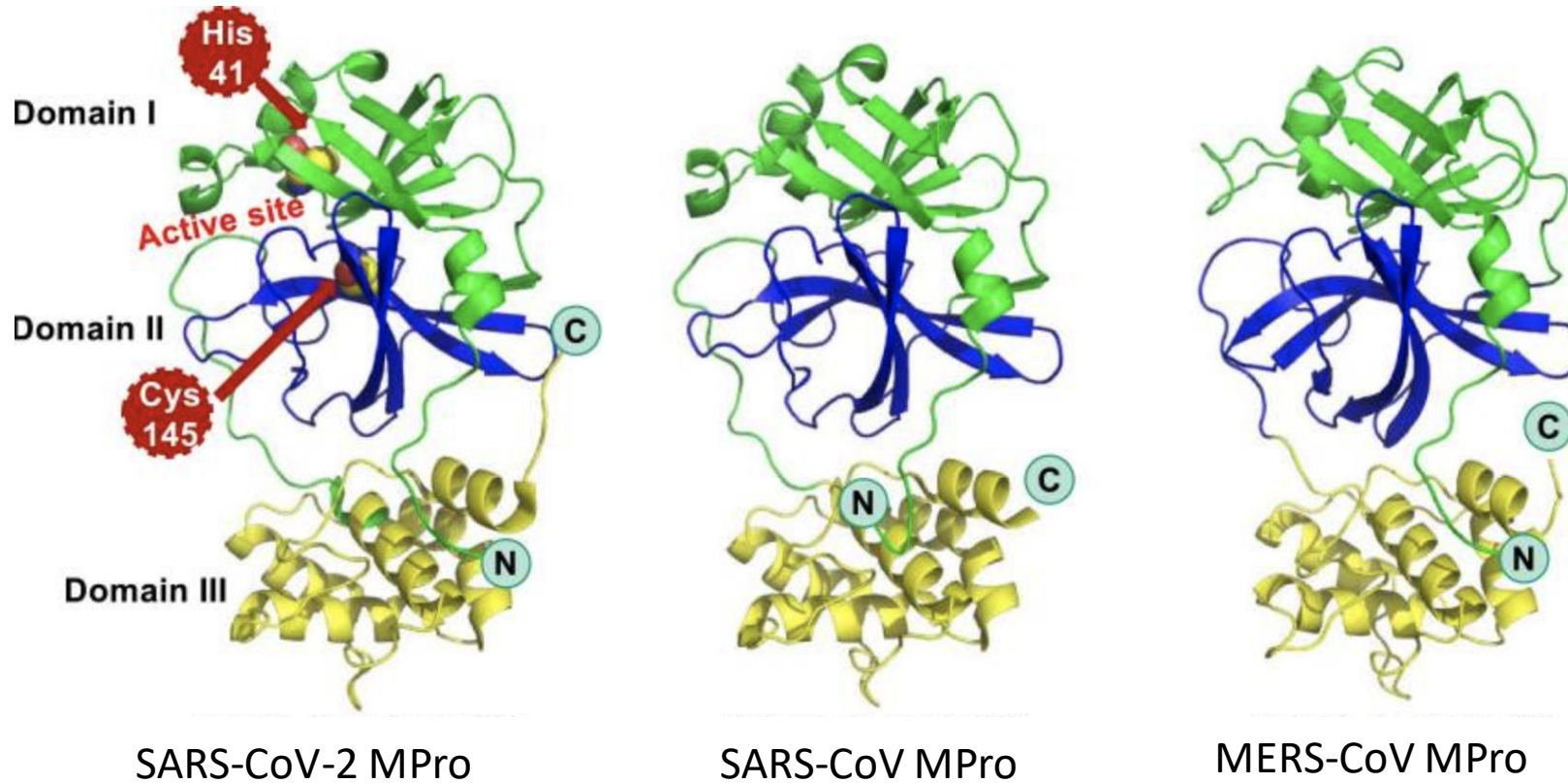


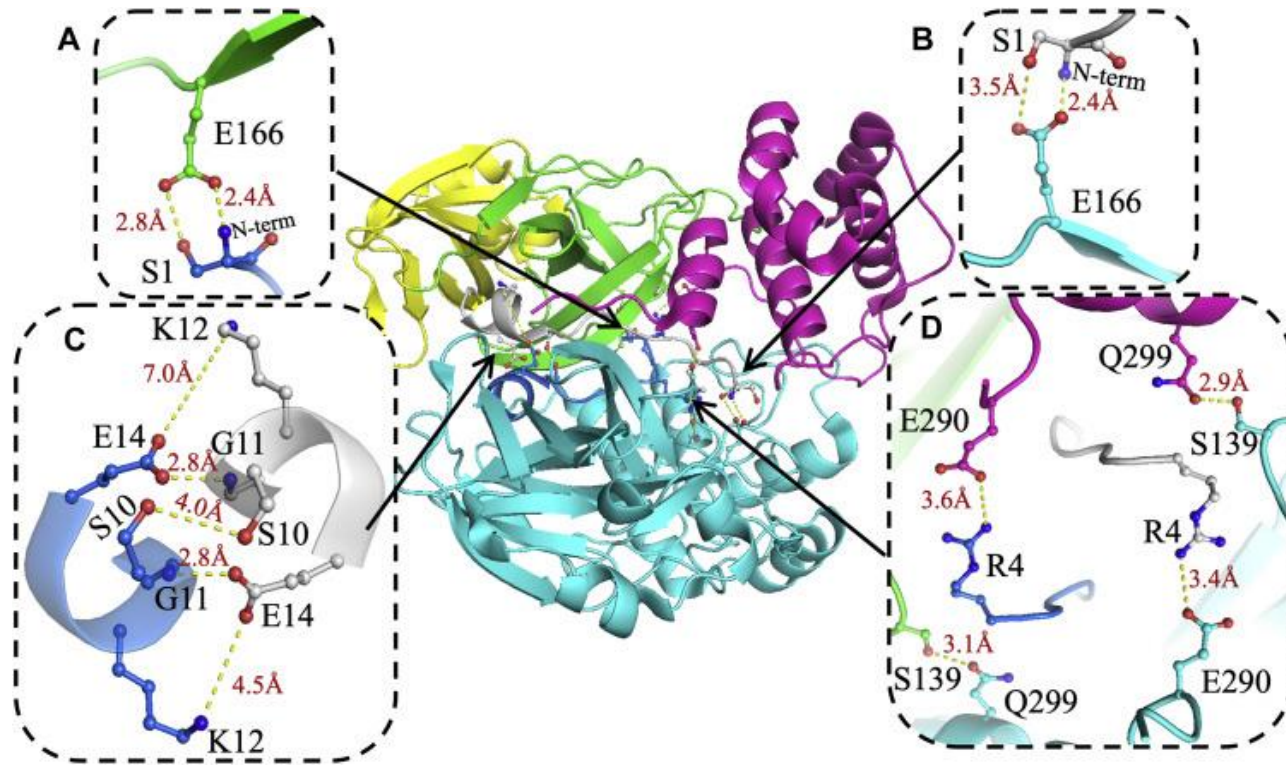
Fig. Three-dimensional structures of SARS-CoV-2 MPro (PDB ID: [6M03](#)), SARS-CoV MPro (PDB ID: [2C3S](#)) and MERS-CoV Mpro (PDB ID: [4YLU](#)). Domains I–III are colored in green, blue and yellow, respectively. Two main amino acid residues (His41 and Cys145) in the catalytic site of SARS-CoV-2 MPro are shown and are colored by atom types.

([Ref](#): Jun He *et al.*, *Int J Antimicrob Agents*, 2020)

SARS-CoV MPro and SARS-CoV-2 MPro: 96% sequence similarity of the 306 residues, only 12 residues are different.

SARS-CoV MPro and MERS-CoV MPro: 51% sequence similarity

Dimerization is essential for catalytic activity of Mpro.

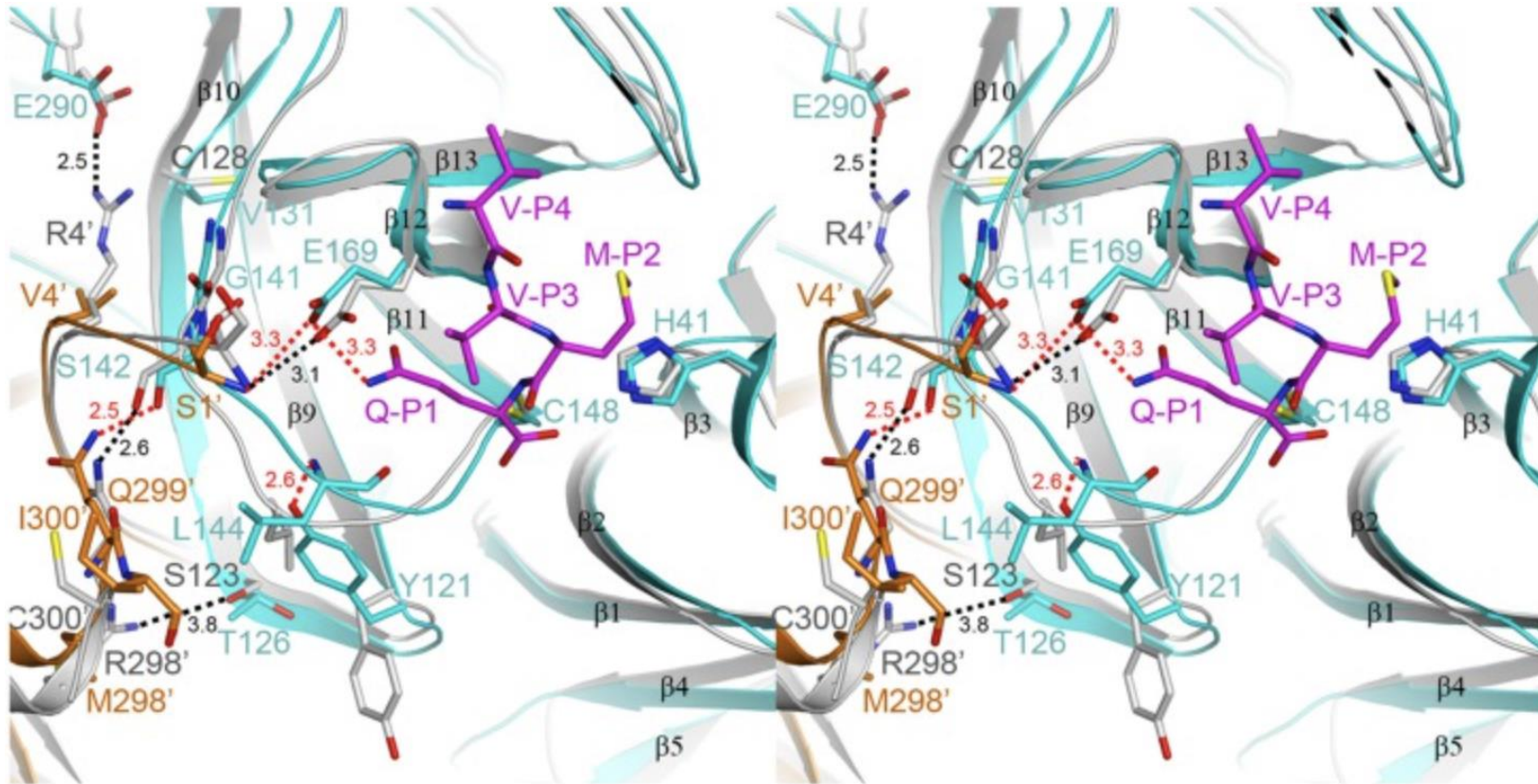


K_d : ranged from $230 \pm 30 \mu\text{M}$ [Ref]
down to $0.19 \pm 0.03 \mu\text{M}$ [Ref]

Fig. Intermolecular interactions at the dimer interface of SARS-CoV-2 Mpro (A and B. Ser1-Glu166; C. Ser10-Ser10, Lys12-Glu14; D. Arg4-Glu290, and Ser139-Gln299)

([Ref](#): Juliana Ferreira *et al.*, *J Biol Chem*, 2022)

Dimerization is essential for catalytic activity of Mpro.



SARS-CoV MPro:

- Four amino acid pairs with intermolecular polar interactions (Ser1-Glu166, Arg4-Glu290, Ser123-Arg298 and Ser139-Gln299).
- K_d : $0.06 \pm 0.01 \mu\text{M}$

MERS-CoV MPro:

- Only two pairs of intermolecular hydrogen bonds (Ser1-Glu169 and Ser142-Gln299)
- K_d : $52 \pm 5 \mu\text{M}$

Fig. Stereo view of an overlay of the dimerization interface and the active site of MERS-CoV MPro (in cyan and orange) with that of SARS-CoV MPro (grey). The red dashed lines indicate polar interactions between the two protomers of MERS-CoV MPro, while the black dashed lines show polar interactions between those of SARS-CoV MPro. The C atoms of the modeled substrate P4-P1 residues (from the structure of C148A mutant) are colored magenta.

(Ref: Bo-Lin Ho *et al.*, *PLoS One*, 2015)

Ligand-induced Dimerization

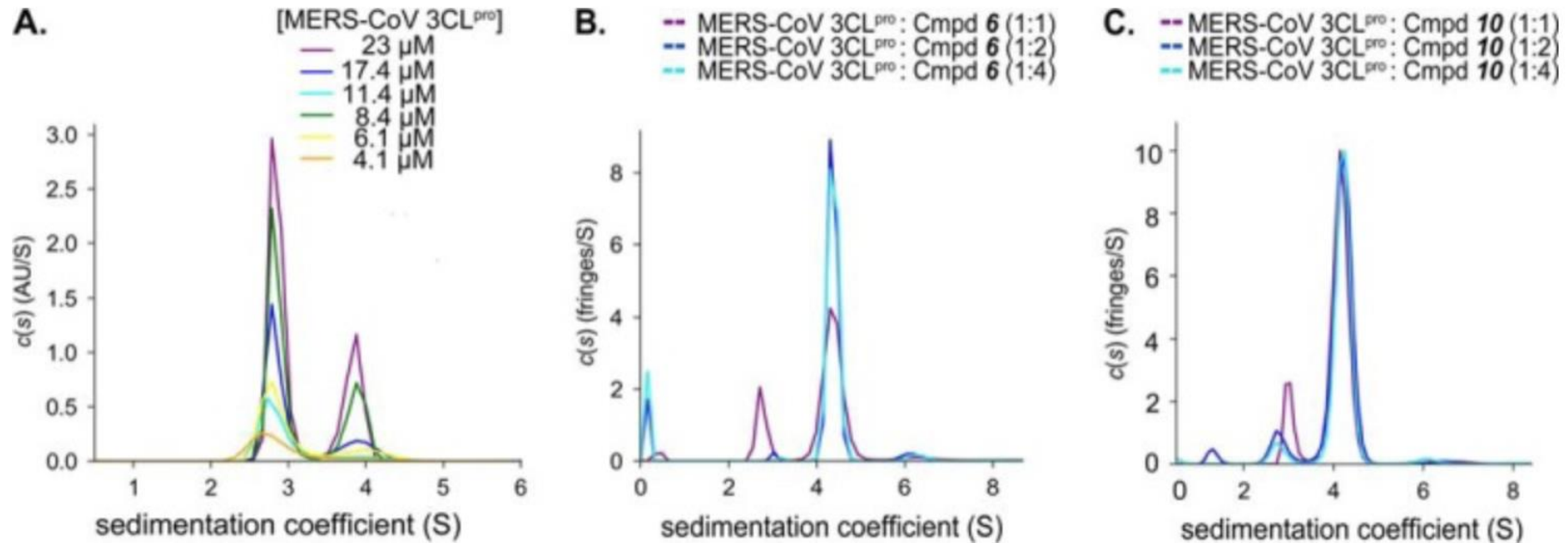


Fig. AUC (Analytical Ultracentrifugation) analyses of ligand-induced dimerization of MERS-CoV MPro.

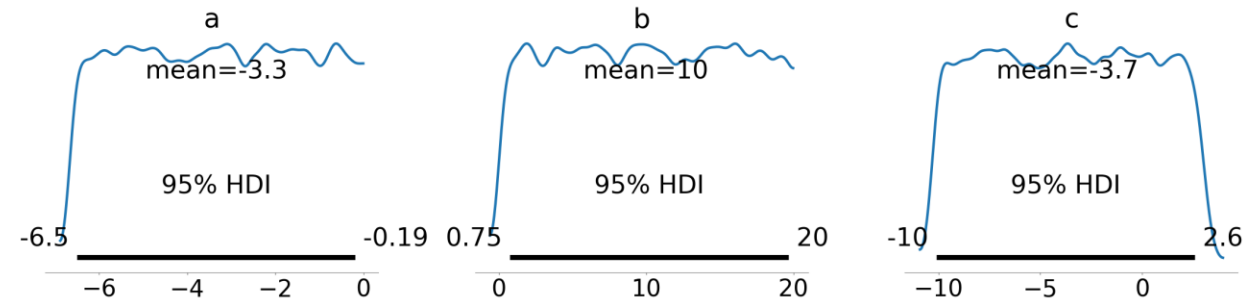
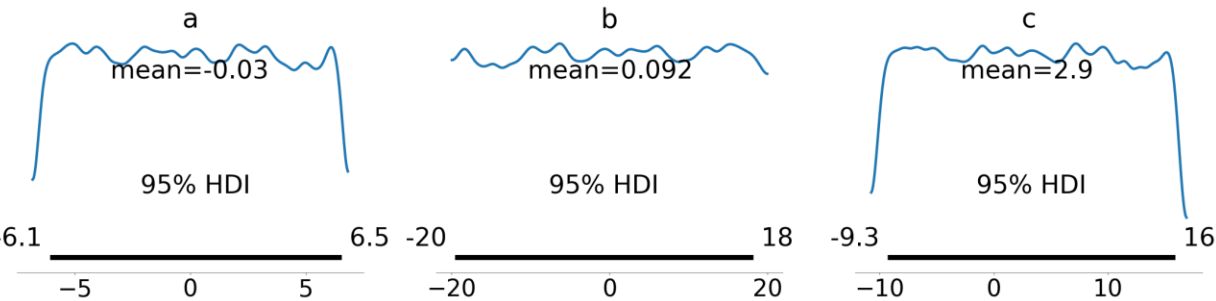
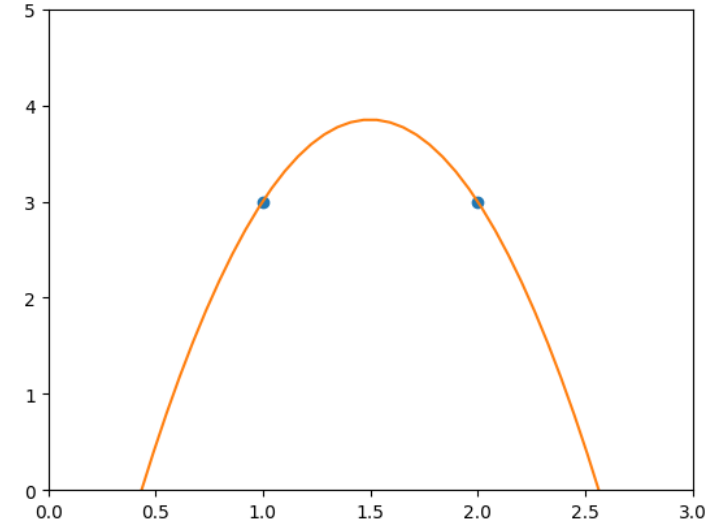
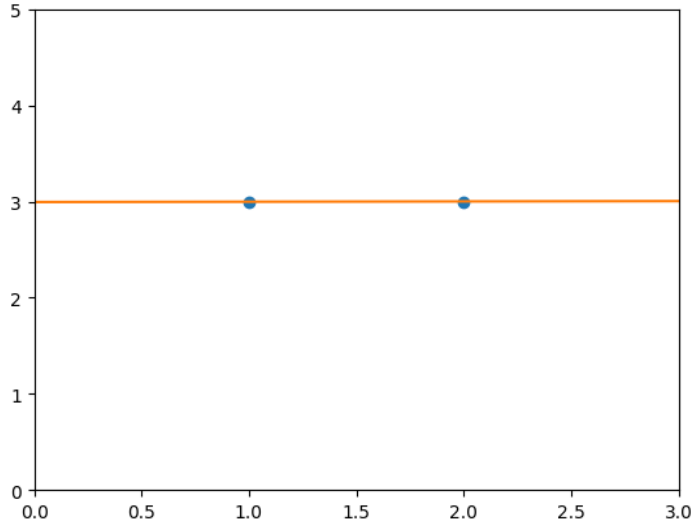
A, sedimentation coefficient distribution for varying concentrations of enzyme (4.1 to 23 μM) with sedimentation coefficient values of 2.9S and 3.9S for the monomer and the dimer, respectively. B, sedimentation coefficient distribution of MERS-CoV 3CL^{pro} (25 μM) in the presence of different stoichiometric ratios of compound **6** (25, 50, and 100 μM). C, sedimentation coefficient distribution of MERS-CoV 3CL^{pro} (25 μM) in the presence of different stoichiometric ratios of compound **10** (25, 50, and 100 μM). A significant shift in the 2.9S peak (monomer) to a 4.1S peak (dimer) is detected upon addition of increasing concentrations of compounds **6** and **10**.

([Ref](#): Sakshi Tomar *et al.*, *J Biol Chem*, 2015)

$$y = ax^2 + bx + c \text{ with } \theta \equiv (a, b, c)$$

Prior: $-20 < a < 20$
 $-20 < b < 20$
 $-20 < c < 20$

Prior: $-20 < a < 0$ (more informative)
 $-20 < b < 20$
 $-20 < c < 20$



(Highest Density Interval - HDI is the interval which contains the required mass such that all points within the interval have a higher probability density than points outside the interval.)

Prior

$$\log K_d^{Mut} \sim \text{Normal}(-6.08, 1.32)$$

$$\log K_{I,D} \sim \text{Normal}(-6.51, 1.32)$$

$$\log K_{I,DI} \sim \text{Normal}(-6.51, 1.32).$$

$$\log K \sim \text{Uniform}(-8.68, 0.00).$$

$$k_{cat}^{Mut} \sim \text{Uniform}(0, 1).$$

$$k_{cat}^{WT} \sim \text{Uniform}(0, 200).$$

$$p(\sigma) \propto \frac{\sigma_0}{\sigma}$$

Likelihood

$$y_n \sim \mathcal{N}(y_n^*(\boldsymbol{\theta}), \sigma^2)$$

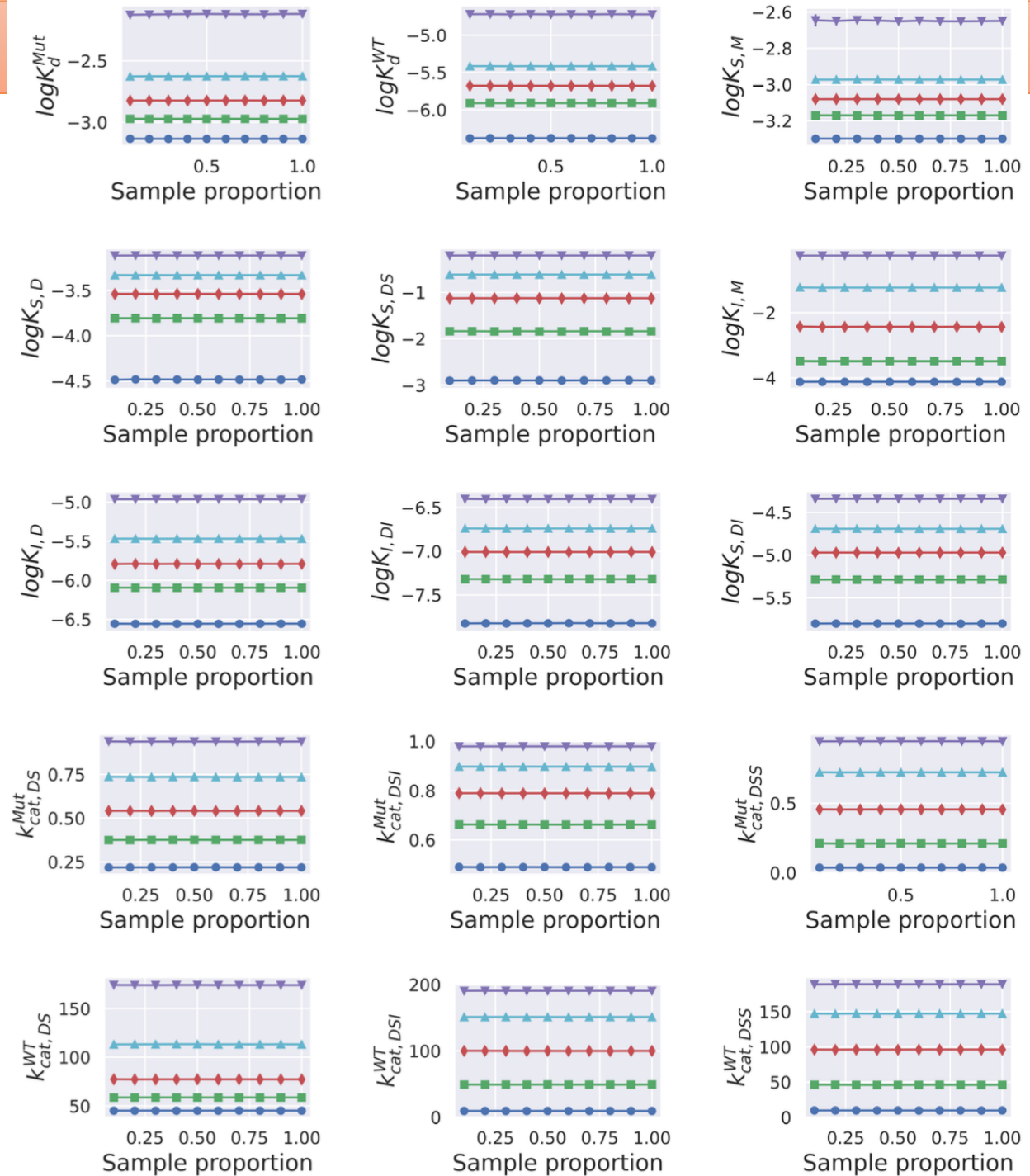
$$p(\mathcal{D}|\boldsymbol{\theta}) = \frac{1}{(2\pi)^{N/2}\sigma^N} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=1}^N (y_n - y_n^*(\boldsymbol{\theta}))^2 \right]$$

Sampling from the posterior

- NUTS sampling
- 4 chains of 2000 warmups, and 10000 samples

AIM 1.1: Results

- Figure. **Convergence of percentiles of the Bayesian posterior.** 10,000 samples were drawn from the Bayesian posterior using the NUTS sampler. All kinetic parameters are shown. Lines correspond to the 5-th (blue circle), 25-th (green square), 50-th (red diamond), 75-th (cyan upward triangle) and 95-th (magenta downward triangle) percentile.



Prior

$$\log K_d \sim \text{Normal}(-5.00, 0.50)$$

$$\log K_{S,M} \sim \text{Uniform}(-9.00, 0.00)$$

$$\log K_{S,D} \sim \text{Uniform}(-9.00, 0.00)$$

$$\log K_{S,DS} \sim \text{Uniform}(-9.00, 0.00)$$

$$\log K_{I,M} \sim \text{Uniform}(-12.00, 0.00)$$

$$\log K_{I,D} \sim \text{Uniform}(-12.00, 0.00)$$

$$\log K_{I,DI} \sim \text{Uniform}(-12.00, 0.00)$$

$$\log K_{S,DI} \sim \text{Uniform}(-12.00, 0.00)$$

$$k_{cat} \sim \text{Uniform}(0.00, 20.00)$$

$$p(\sigma) \propto \frac{\sigma_0}{\sigma} \quad \alpha^i \sim \text{Uniform}(0, 2)$$

$$\log[E]^k \sim \mathcal{N}(\mu = [E]_0^k, \sigma = 0.1 * [E]_0^k)$$

Likelihood

$$y_n \sim \mathcal{N}(\alpha y_n^*(\boldsymbol{\theta}), \sigma^2)$$

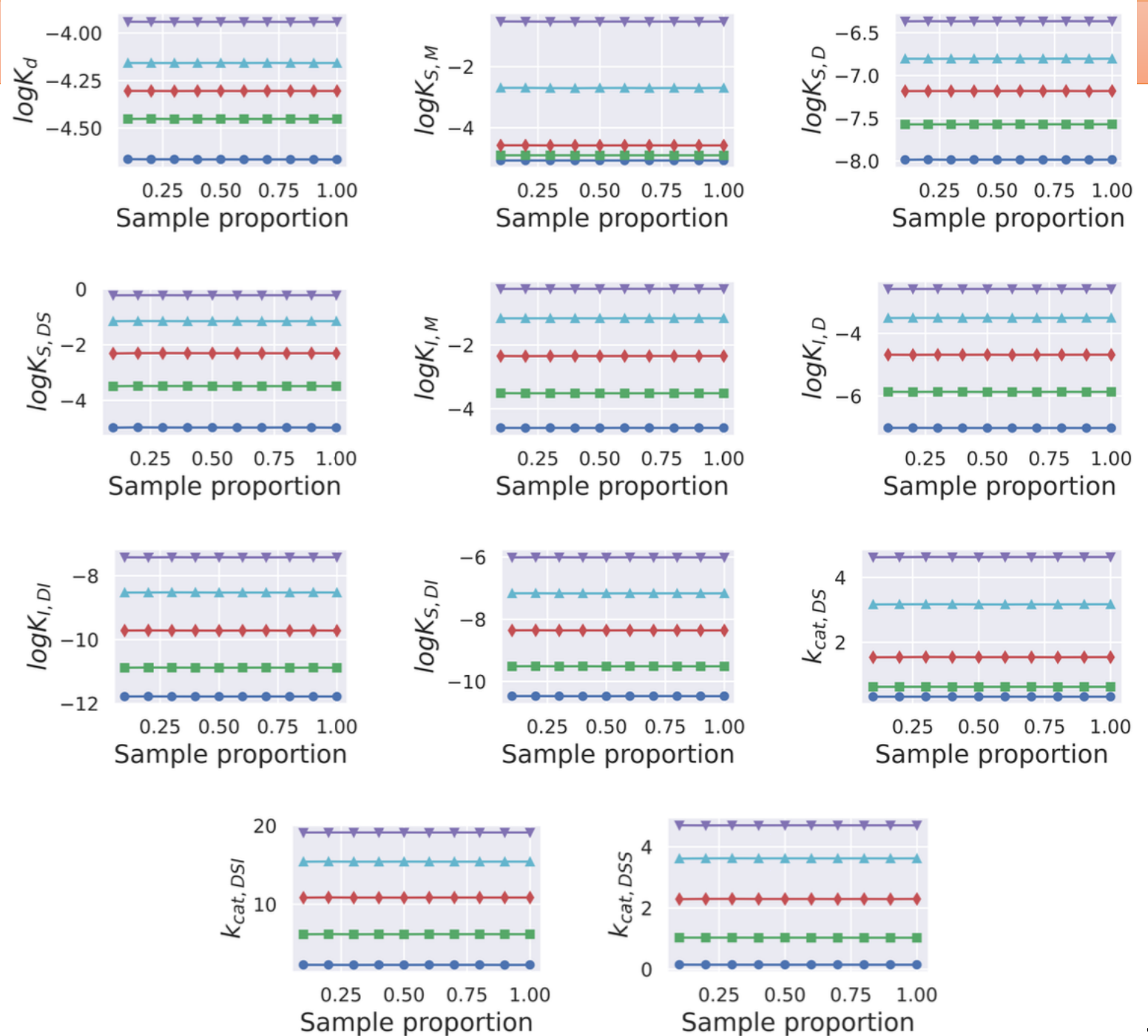
$$p(\mathcal{D}|\boldsymbol{\theta}) = \frac{1}{(2\pi)^{N/2} \sigma^N} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=1}^N (y_n - \alpha y_n^*(\boldsymbol{\theta}))^2 \right]$$

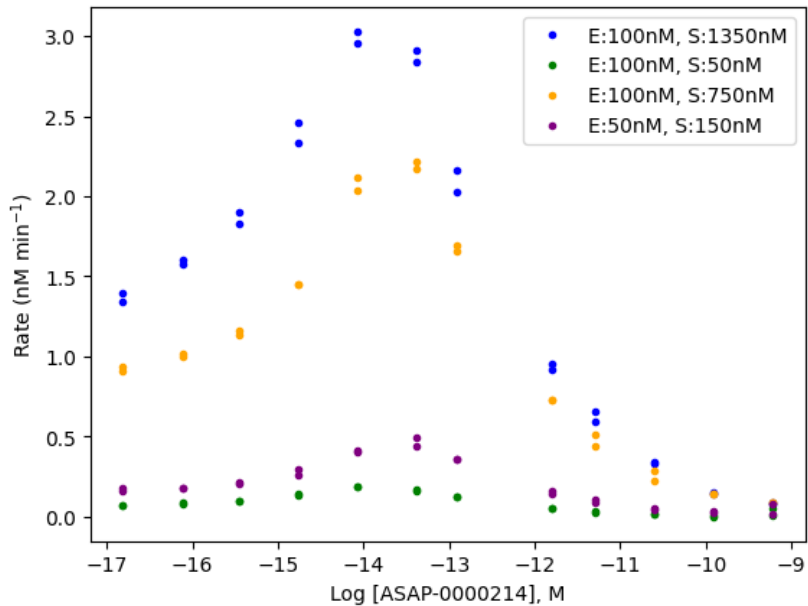
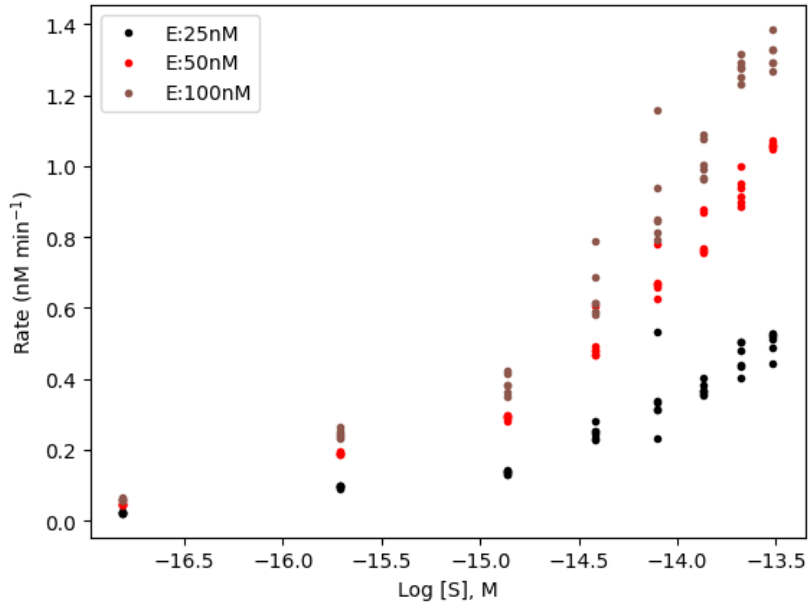
Sampling from the posterior

- NUTS sampling
- 4 chains of 2000 warmups, and 10000 samples

AIM 2.1: Results

- Figure. **Convergence of percentiles of the Bayesian posterior of dissociation and rate constants.** 10,000 samples were drawn from the Bayesian posterior using the NUTS sampler. All kinetic parameters are shown. Lines correspond to the 5-th (blue circle), 25-th (green square), 50-th (red diamond), 75-th (cyan upward triangle) and 95-th (magenta downward triangle) percentile.





Set of parameters

$$\theta \equiv (K_d, K_{S,M}, K_{S,D}, K_{S,DS}, K_{I,M}, K_{I,D}, K_{I,DI}, K_{I,DI}, K_{S,DI}, k_{cat,MS}, k_{cat,DS}, k_{cat,DSI}, k_{cat,DSS}, \alpha^i, [E]^k, \sigma^j).$$

- K and k_{cat} are global parameters.
- α^i is normalizing factor for each plate
- σ_j is assumed to be constant for the dataset j^{th} .
- $[E]^k$ with $k = \{100, 50, 25\}$ (unit: nM).

Model fitted datasets.

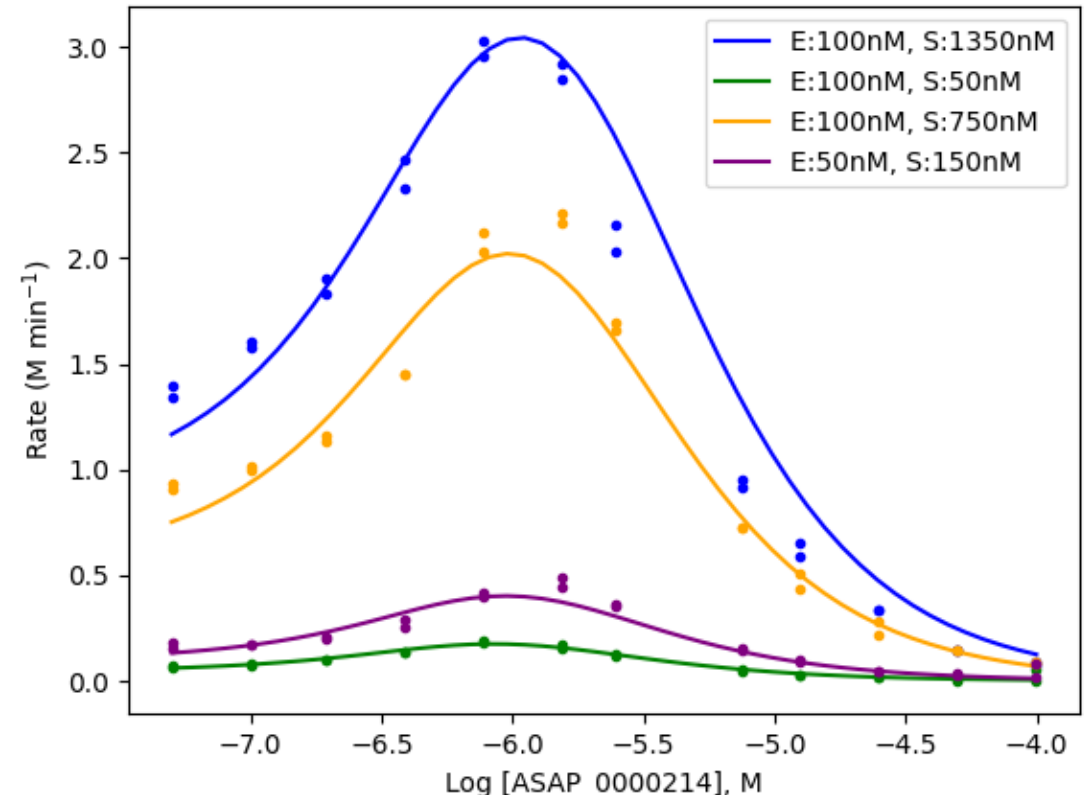
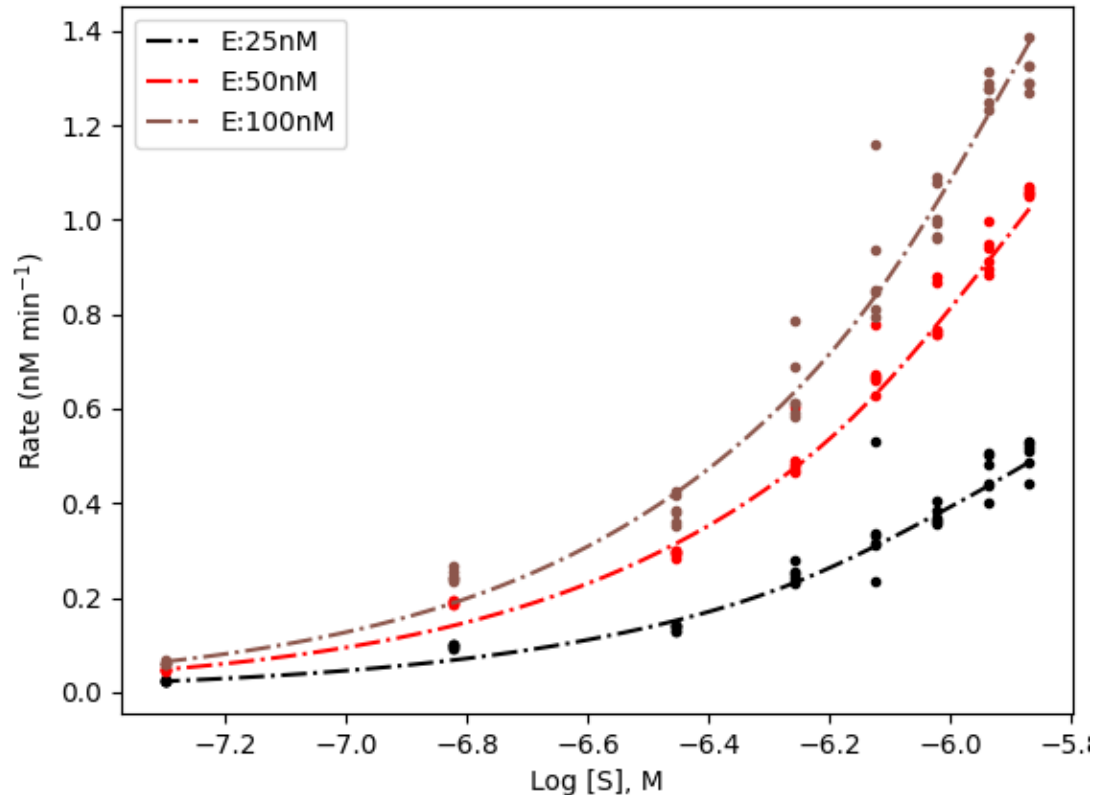


Figure. MAP fitted ES and ASAP-0000214 datasets

The lines are the theoretical responses $y_n * (\theta_{\text{MAP}})$, where θ_{MAP} is the Maximum a Posteriori estimate of the parameters. Dots are the observed response.

Aim 1: To construct the Bayesian model for the estimation of kinetic parameters for one set of inhibitor

1.1. Fitting the Bayesian model from SARS-CoV-2 MPro datasets

1.2. Simplifying model by adding the constraints on parameters

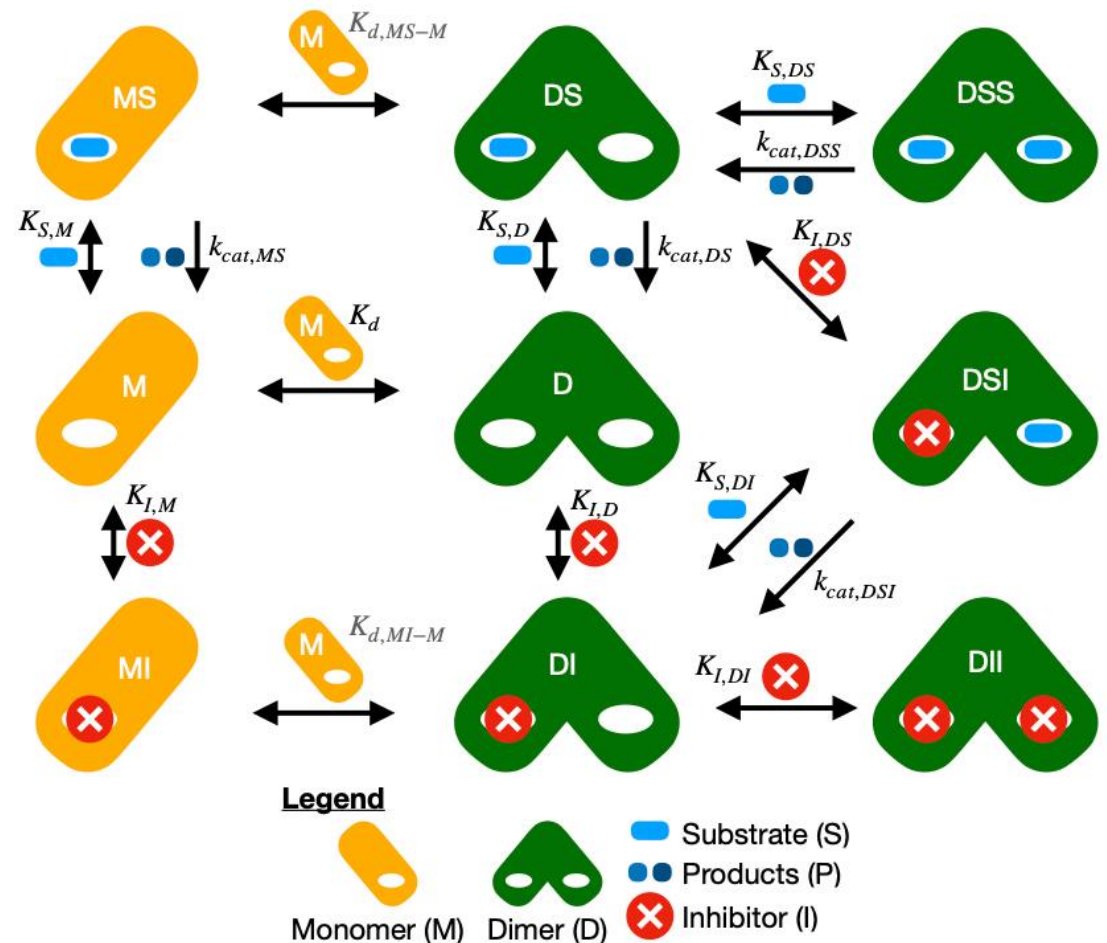
AIM 1.2: Methods

Based on model result:

- $K_{S,M} = K_{I,M}$
- $K_{S,DI} = K_{S,DS}$
- $k_{cat_{DS}} = k_{cat_{DSS}}$
- $k_{cat_{DS}} = k_{cat_{DSI}}$
- $k_{cat_{DSI}} = k_{cat_{DSS}}$

The symmetry of model:

- $\log K_{S,D} - \log K_{S,M} = \log K_{I,D} - \log K_{I,M}$
- $\log K_{I,DS} - \log K_{S,DS} = \log K_{S,DI} - \log K_{I,DI}$

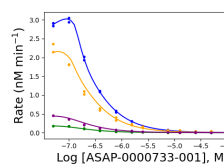
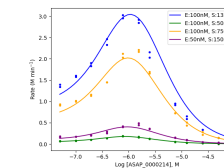
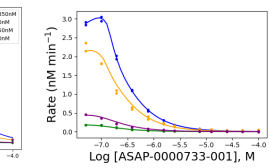
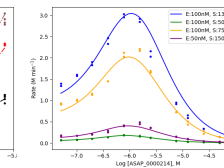
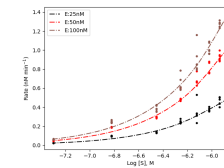
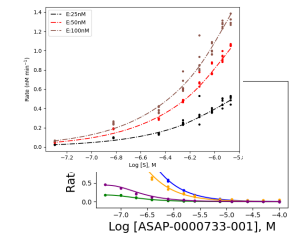
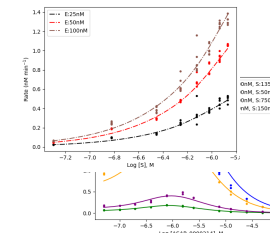
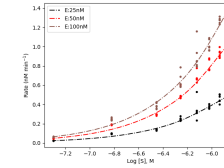


AIM 1.2: Methods

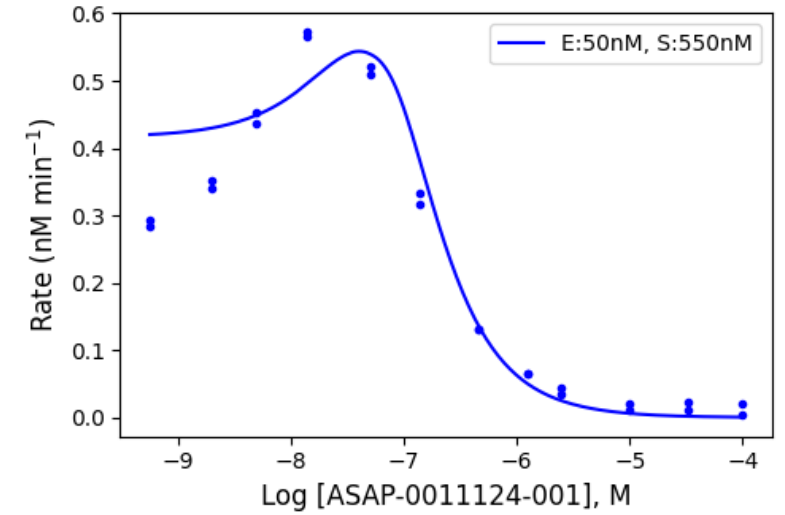
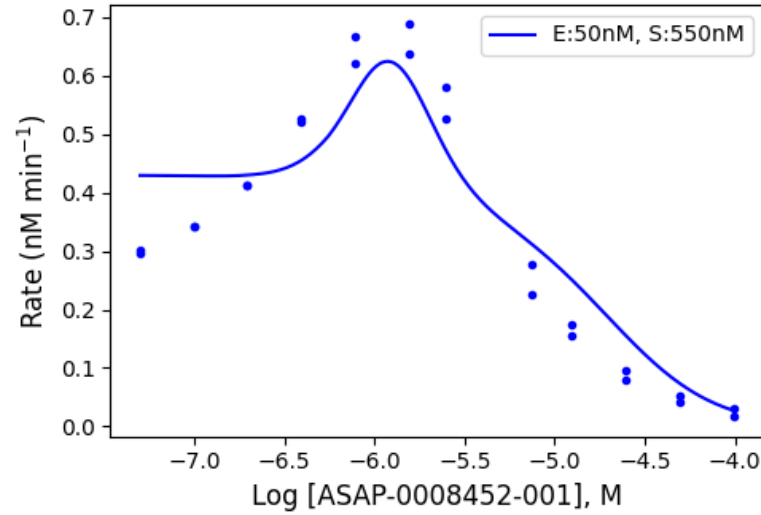
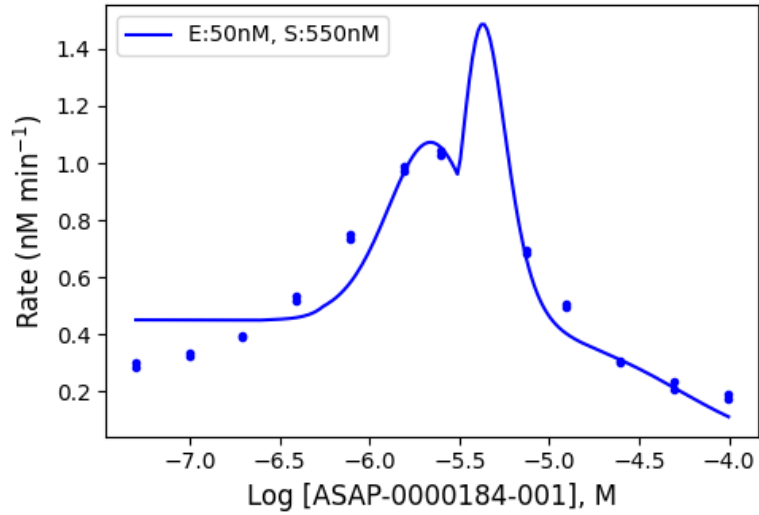
	Model convergence	MAP fitted all datasets
$K_{S,M} = K_{I,M}$		
$K_{S,DI} = K_{S,DS}$		
$kcat_{DS} = kcat_{DSS}$	x	
$kcat_{DS} = kcat_{DSI}$	x	
$kcat_{DSI} = kcat_{DSS}$	x	
$\log K_{S,D} - \log K_{S,M} = \log K_{I,D} - \log K_{I,M}$		
$\log K_{I,DS} - \log K_{S,DS} = \log K_{S,DI} - \log K_{I,DI}$	x	

Global fitting procedure

1. We fitted only ES datasets and estimated $\log K_d$, $\log K_{S,M}$, $\log K_{S,D}$, $\log K_{S,DS}$, $k_{cat,DS}$, $k_{cat,DSS}$, and $[E]$.
 α was set at 1 for this dataset.
2. The posterior distributions of parameters obtained in step 1 were used to adjust the prior of those parameters for the separate fitting of each ESI dataset. Different α and $[E]$ parameters were assigned for different plates.
3. The overlapping range of shared parameters $\log K$, k_{cat} , α , and $[E]$ were extracted and used as the prior information for the global fitting of all datasets.
4. Values of shared parameters were fixed based on the MAP of the posterior distribution from step 3, and Bayesian model is fitted once more for ESI.



Fitting each CRC



Extended global fitting for a subset of data



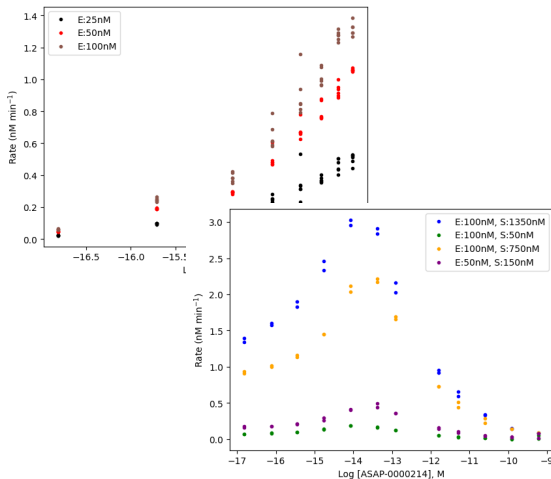
Fitting each CRC



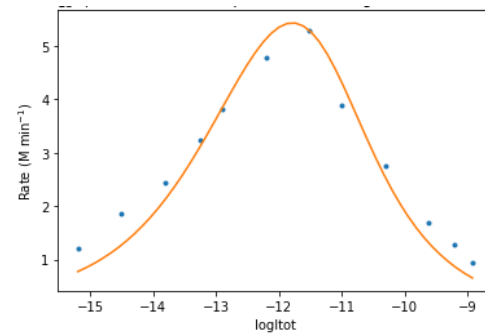
Dimer-only pIC50 for each CRC



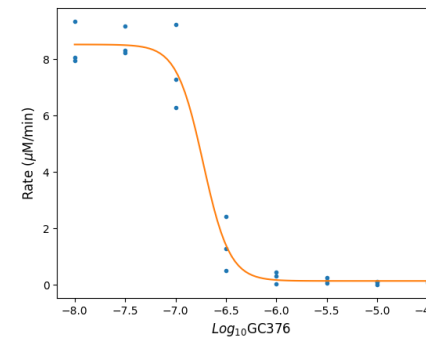
Correlation analysis



1 set of ES and + 15 sets of ESI (fitting 157 parameters)



Fitting 8 parameters for each CRC



Log[E] ~ N(μ=50.0, σ=2.5)
Log[S] ~ N(μ=1350.0, σ=67.5)
(unit: nM)

$$y_i = R_{base} + \frac{R_{max} - R_{base}}{1 + 10^{(\log Concentration_i - \log IC_{50}) * hill}}$$

- Pearson R
- Spearman ρ
- Kendall τ
- RMSD
- aRMSD

Extended global fitting: Results

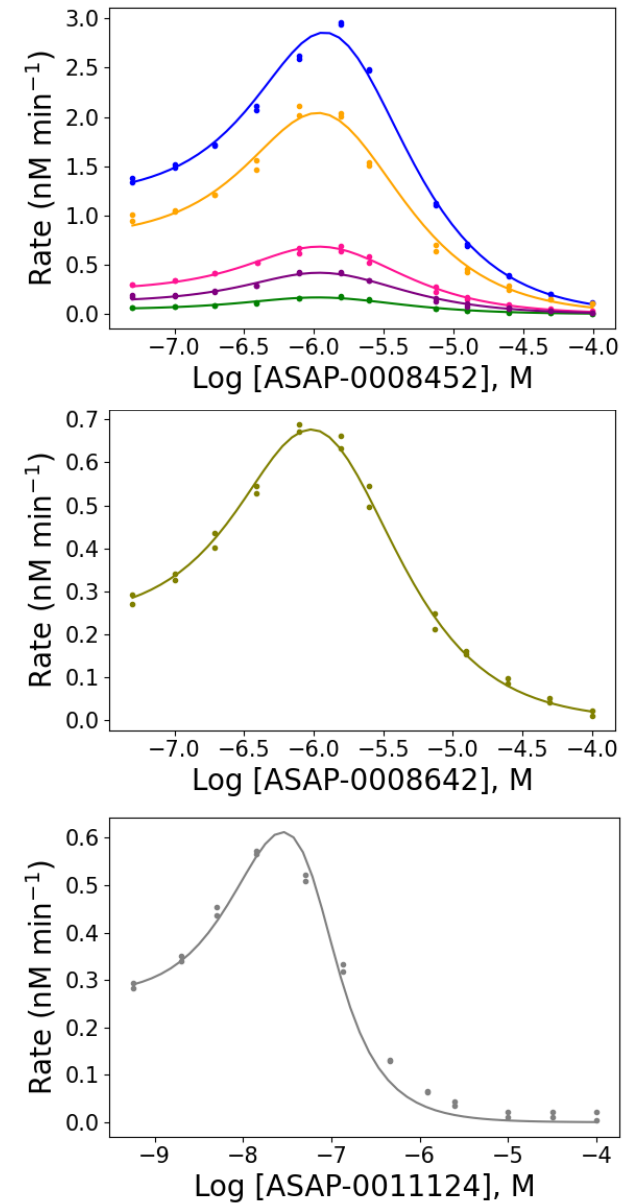
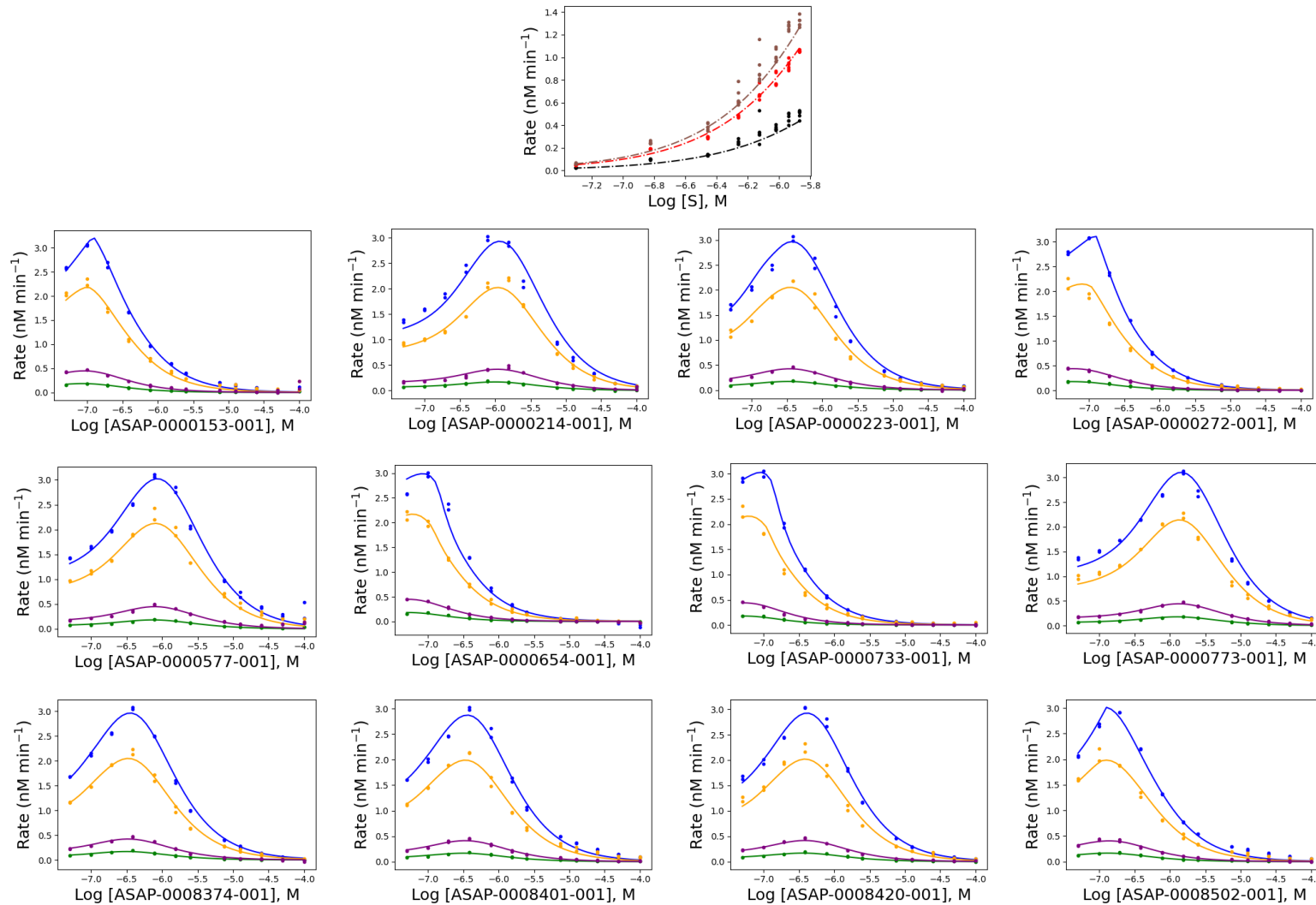


Figure. MAP fitted ES and ESI inhibitor datasets

The lines are the theoretical responses $y_n * (\theta_{MAP})$. Dots are the observed response. Curves in the same colors represent for the datasets in the same plates.