Supporting Information:
Bayesian Data Integration Questions Classic
Study on Protease Self-digest Kinetics

Niklas Tötsch* and Daniel Hoffmann

Bioinformatics and Computational Biophysics, Universität Duisburg-Essen, 45141 Essen, Germany

E-mail: niklas.toetsch@uni-due.de
S1 Pre-processing of data from JW1981

Due to the considerable time that has passed since JW1981 was published, we had to access their data by digitizing their Figures 4 and 8.\(^1\)

JW1981 did not measure the concentration of trypsin directly but spectroscopically monitored the tryptic digest of substrate benzoyl-L-arginine ethyl ester (BAEE). If the substrate is present in great excess, the concentration of trypsin limits the reaction rate. The concentrations were then determined by comparison of the observed rates with corresponding data from experiments with known trypsin concentrations.

For the kinetic experiments, data in JW1981 were given as the inverse of relative activity. We obtained relative activities by inverting these values. JW1981 noted that adsorption of trypsin on silica does not change its activity. Overall activity is therefore the sum of \([T]\) and \([TS]\) (see Section S5.6). Multiplying relative activities by the known starting concentrations of trypsin yields \([T]_t + [TS]_t\).

The amount of adsorbed trypsin in Figure 8 in JW1981 is given in \(\text{mg trypsin} / \text{mg } \text{SiO}_2\). We converted these data by dividing by trypsin’s molecular weight (23.3 kg mol\(^{-1}\)) and multiplying with the amount of silica present in the kinetic experiments (10 mg l\(^{-1}\)) to arrive at the concentration of \([TS]\) in mol l\(^{-1}\).

S2 Mechanism of surface accelerated self-digest

Colloidal silica in water has a disperse phase. There are two major mechanisms commonly used to describe surface reactions (SRs) in heterogeneous catalysis, Eley-Rideal (ER) and Langmuir-Hinshelwood (LH). Both mechanisms appear in textbooks for reactions with two different reactants and here have been slightly adapted to cater to trypsin-trypsin digest with two identical reactants.
S2.1 surface coverage

Trypsin molecules can bind to and unbind from silica particle surfaces.

\[
T + S \xrightleftharpoons[k_d]{k_a} TS \tag{S1}
\]

where:

- \(S\) = unoccupied binding site on the silica surface
- \(TS\) = trypsin adsorbed on the silica surface
- \(k_a\) = association rate
- \(k_d\) = dissociation rate

Koutsopoulos et al. estimated that a tightly packed monolayer of trypsins would lead to a coverage of 2.1 mg m\(^{-2}\). On silica slides, they observed a maximum coverage of 1.1 mg m\(^{-2}\) whereas Berg et al. observed a peak around 2 mg m\(^{-2}\). Syton X30, the colloidal silica used by JW1981, has a specific surface of 242 m\(^2\) g\(^{-1}\). Multiplied by Koutsopoulos’ estimate of the density of a tightly packed monolayer, Syton X30 could bind up to 0.51 \(\frac{mg\ T}{mg\ SiO_2}\). The adsorption isotherm presented in JW1981 appears to converge to a similar value.

While the experiments by Berg and JW1981 apparently agree on maximum coverage, they suggest different values for the binding constant \(K_a = \frac{k_a}{k_d}\). Berg’s experiments approach full monolayer coverage, i.e. \(\phi \approx 1\), with an initial trypsin concentration \([T]_0 = 0.43 \mu M\). This implies a larger \(K_a\) than what we found for the adsorption isotherm (\(D_{ADS}\)) from JW1981. JW1981’s adsorption experiments were performed in 50 mM NaCl, 2 mM CaCl\(_2\), 10 mM Tris/HCl buffer at pH 8.0 and colloidal silica, whereas Berg used 10 mM phosphate buffer at pH 7.6 and silica wafers. If the Ca\(^{2+}\) ions were responsible for this difference in \(K_a\), JW1981’s data sets \(D_{ADS}\) and surface accelerated self-digest (\(D_{SAS}\)) (measured in absence of Ca\(^{2+}\)) could not be integrated in a consistent Bayesian inference.

\(^1\)https://www.kremer-pigmente.com/media/pdf/31430.pdf accessed on October 6th 2019 at 11:53.
S2.2  Eley-Rideal (ER)

In 1938, Eley and Rideal proposed a mechanism where an adsorbed molecule reacts with an unbound partner (Scheme S1).

\[
TS + T \xrightarrow{k_{ER}} T + S + P
\]  

(S2)

The reaction rate on the surface \( r_{\text{ER}}^{\text{surface}} \) (in \( \text{mol} \, \text{m}^{-2} \, \text{s}^{-1} \)) equals

\[
r_{\text{ER}}^{\text{surface}} = \left( \frac{d[P]^{\text{surface}}}{dt} \right)_{\text{ER}} = C_S k_{ER}^{\prime\prime} \phi[T]
\]  

(S3)

where:

\( C_S \) = binding site density on the surface, \( \text{mol} \, \text{m}^{-2} \)

\( k_{ER}^{\prime\prime} \) = rate constant, \( \text{m}^{-1} \, \text{s}^{-1} \)

The binding site density is dependent on the size of the adsorbent and the properties of the surface. Since we are only interested in one pair of adsorbent and surface, we do not model \( C_S \) explicitly. The reaction rate for the system \( r_{\text{ER}}^{\text{solution}} \) (in \( \text{M}^{-1} \, \text{s}^{-1} \)) is defined by
\[ r_{\text{solution}}^{\text{ER}} = \left( \frac{d[P]^{\text{solution}}}{dt} \right)_{\text{ER}} = A r_{\text{surface}}^{\text{ER}} = AC S k_{\text{ER}}'' \phi[T] = k_{\text{ER}}' \frac{[TS]}{[S]_0} [T] = k_{\text{ER}} [TS][T] \]  

(S4)

where:

\[ A \quad = \text{silica surface per volume, m}^2 \text{ L}^{-1} \]
\[ k_{\text{ER}}' = AC S k_{\text{ER}}'' \]
\[ k_{\text{ER}} = \frac{k_{\text{ER}}'}{[S]_0} \]

We do not need to model A explicitly because JW1981 always worked with the same concentration of silica. The complete system of ordinary differential equations is

\[
\begin{align*}
\frac{d[P]}{dt} &= k_{\text{BR}}[T]^2 + k_{\text{ER}}[T][TS] \quad \text{(S5)} \\
\frac{d[S]}{dt} &= -k_a[S][T] + k_d[TS] + k_{\text{ER}}[T][TS] \quad \text{(S6)} \\
\frac{d[T]}{dt} &= -k_{\text{BR}}[T]^2 - k_a[S][T] + k_d[TS] \quad \text{(S7)} \\
\frac{d[TS]}{dt} &= k_a[S][T] - k_d[TS] - k_{\text{ER}}[T][TS]. \quad \text{(S8)}
\end{align*}
\]

If unbound and adsorbed trypsin are in equilibrium, the reaction order (\( \omega \)) w.r.t. to [T] can be determined from

\[
\omega_{\text{ER}}^{[T]} = \frac{d \ln r_{\text{solution}}^{\text{ER}}}{d \ln[T]} = 2 - \phi. \quad \text{(S9)}
\]

If surface coverage (\( \phi \)) can be considered constant, the rate of product formation is dependent only on the concentration of unbound trypsin in solution. If \( \phi \) is close to 1, i.e. maximum coverage, the reaction is first-order.

If \( \phi \) is approximately 0, product formation resembles a second-order reaction. Adding
trypsin to the system results in an increase in adsorbed trypsin. For intermediate degrees of surface coverage, the observed reaction would appear intermediate between first- and second-order.

Eq. S2 assumes that the substrate is adsorbed, whereas the digester attacks from the bulk. In principle, these roles could also be reversed with an immobilized digester on the surface proteolysing substrates from the solution:

\[ TS + T \xrightarrow{k_{ER}} TS + P \]  

(S10)

No trypsin adsorbed on the silica surface (TS) is consumed during autolysis in this model. We did not consider this variant of the ER mechanism. Surface reaction (SR) must be faster than bulk reaction (BR) in \( D_{SAS} \). There is no obvious reason why adsorbed enzyme should proteolyse substrate from solution faster than unbound enzyme. JW1981 found no difference in enzymatic activity towards BAEE upon adsorption. Moreover, this mechanism would most likely have the same shortcomings as JW1981’s hypothesis discussed in the main text.

### S2.3 Langmuir-Hinshelwood (LH)

Langmuir and Hinshelwood developed a mechanism where both molecules are adsorbed on the surface before the reaction (Scheme S2).
\[ 2 \text{TS} \xrightarrow{k_{LH}''} \text{TS} + \text{P} + \text{S} \quad (S11) \]

\[ r_{LH}^\text{surface} = \left( \frac{d[P]^\text{surface}}{dt} \right)_{LH} = C_S^2 k_{LH}'' \phi \phi \quad (S12) \]

where:

\[ k_{LH}'' = \text{rate constant, m}^2 \text{M}^{-1} \text{s}^{-1} \]

\[ r_{LH}^\text{solution} = A r_{LH}^\text{surface} = \left( \frac{d[P]^\text{solution}}{dt} \right)_{LH} = AC_S^2 k_{LH}'' \phi \phi = k_{LH}' \phi \phi = k_{LH}[\text{TS}][\text{TS}] \quad (S13) \]

where:

\[ k_{LH}' = AC_S^2 k_{LH}'' \]

\[ k_{LH} = \frac{k_{LH}'}{[S_0]^2} \]

Contrary to ER, the rate of product formation does not depend directly on the concentration of unbound trypsin. The complete set of differential equations is

\[ \frac{d[P]}{dt} = k_{BR}[T]^2 + k_{LH}[\text{TS}]^2 \quad (S14) \]

\[ \frac{d[S]}{dt} = -k_a[S][T] + k_d[\text{TS}] + k_{LH}[\text{TS}]^2 \quad (S15) \]

\[ \frac{d[T]}{dt} = -k_{BR}[T]^2 - k_a[S][T] + k_d[\text{TS}] \quad (S16) \]

\[ \frac{d[TS]}{dt} = k_a[S][T] - k_d[\text{TS}] - k_{LH}[\text{TS}]^2. \quad (S17) \]

If unbound and adsorbed trypsin are in equilibrium, the reaction order \((\omega)\) equals
\[ \omega_{LH}^{T} = \frac{d \ln r_{LH}^{\text{solution}}}{d \ln[T]} = 2 \cdot (1 - \phi). \] (S18)

If \( \phi \) approaches maximum coverage, product formation converges to a maximum since the maximum amount of available [TS] is capped by the number of binding sites \([S]_0\). The reaction appears zeroth order with respect to \([T]\). For \( \phi \approx 0 \), the reaction appears to be second-order. If \( \phi \approx 0.5 \), the reaction appears to be first-order.

**S2.4  Final model: ER-like surface reaction with two trypsin conformations**

\[ \frac{d[A]}{dt} = -k_{AB}[A] + k_{BA}[B] - k_a[A][S] + k_d[AS] \] (S19)

\[ \frac{d[AS]}{dt} = k_a[A][S] - k_d[AS] \] (S20)

\[ \frac{d[B]}{dt} = k_{BA}[A] - k_{BR}[A][B] - k_a[B][S] + k_d[BS] \] (S21)

\[ \frac{d[BS]}{dt} = k_a[B][S] - k_d[BS] - k_{ER}[A][BS] \] (S22)

\[ \frac{d[P]}{dt} = k_{BR}[A][B] + k_{ER}[A][BS] \] (S23)

\[ \frac{d[S]}{dt} = -k_a[S]([A] + [B]) + k_d([AS] + [BS]) + k_{ER}[A][BS] \] (S24)

**S3  Requirement for constant surface coverage**

We continue to investigate how large the binding constant \( K_a \) would need to be to yield constant surface coverage for the trypsin starting concentrations of interest, i.e. satisfy

\[ \phi_c = \frac{\phi([T]_0^a)}{\phi([T]_0^b)} \approx 1. \] (S25)

Setting \([T]_0^a = 11.0 \mu M \) and \([T]_0^b = 2.2 \mu M \), \( K_a \) needs to be approximately one order of magnitude larger than inferred from \( D_{ADS} \) for \( \phi_c < 1.05 \) (Figure S1). \( \phi_c = 1.05 \) implies a 5%
difference in surface coverage in the beginning for the two starting concentrations of trypsin. There are some differences between the $D_{SAS}$ and $D_{ADS}$ experiments which might lead to a distortion in $K_a$ (see main text).

Figure S1: $K_a$ would need to be substantially larger than implied by $D_{ADS}$ for constant $\phi$.

S4 Regular Trypsin Autolysis

S4.1 Single trypsin conformation

We describe the bulk reaction (BR) of trypsin autolysis with a single bimolecular reaction, neglecting the reversible formation of the digester-substrate complex included in the Michaelis-Menten model.

$$2 \ T \xrightarrow{k_{BR}} T + P \quad (S26)$$

where:

$T$ = unbound trypsin

$P$ = degradation product of trypsin

$k_{BR}$ = rate constant
S4.2 Two trypsin conformations

Assuming two states A, B for trypsin, a simplified version of the autolysis reaction as published by e.g. Nord but neglecting the digester-substrate complex is given by

\[ A \xrightleftharpoons[k_{BA}]^{k_{AB}} B \quad (S27) \]
\[ A + B \xrightarrow{k_{RR}} A + P \quad (S28) \]

S4.3 Reaction order of regular autolysis is below 2

One can approximate the reaction order (\( \omega \)) w.r.t. to \([T]\) by plotting the logarithm of the reaction rate \( r \) against the logarithm of \([T]\) and calculating the slope:

\[ \log r = \log \frac{d[P]}{dt} = \omega \cdot \log([T]) + C \quad (S29) \]

First, we calculate

\[ r = \frac{[T]_{t_2} - [T]_{t_1}}{t_2 - t_1} \quad (S30) \]

from data on regular self-digest (\( D_{REG} \)) for the smallest timestep \( t_2 - t_1 \) available. We then perform Bayesian inference of \( \omega \) and \( C \) (which is a constant of no further interest to us) with this probabilistic model:
\[
\log r^{\text{exp}} \sim N(\log r^{\text{model}}, \sigma)
\]
\[
\log r^{\text{model}} \sim \omega \cdot \log([T]) + C
\]
\[
\omega \sim \text{unif}(0.0, 3.0) \quad (S31)
\]
\[
C \sim \text{unif}(-20.0, 20.0)
\]
\[
\sigma \sim \text{unif}(0.001, 10.0)
\]

Figure S2: Logarithm of reaction rates \(r\) plotted against logarithms of trypsin concentrations in \(D_{\text{REG}}\) and posterior predictive check (left). Marginalized posterior distribution for \(\omega\) (right).

**S4.3.1 Inclusion of two trypsin states can explain \(D_{\text{REG}}\) reaction-order**

Slow conversion of trypsin state A to state B is a likely explanation why the reaction order in JW1981 appears to be below 2 (see main text). Including these two forms improves reproduction of \(D_{\text{REG}}\) (Figure S3).
Figure S3: Posterior predictive check for two models of regular autolysis.

S4.4 Reaction rates in $D_{\text{REG}}$ and $D_{\text{SAS}}$

Reaction rates with and without silica differ strongly for $[T]_0 = 2.2 \ \mu M$ but not for $[T]_0 = 11.0 \ \mu M$ (Figure S4).

S5 Implicit assumptions

In addition to the explicit assumptions that are included in the chemical equations (e.g. Schemes 1a, 1b, 2 in the main text), there are several implicit assumptions:

S5.1 The degradation product (P) does not occupy binding site on the silica surface (S)

An intact protease binds multivalently to the surface. Degradation product (P) consists of smaller fragments which in general cannot engage in multivalent interactions. Berg et al. monitored the surface coverage of a silica surface in a solution of trypsin over time. At first,
Figure S4: Experimental reaction rates for different trypsin starting concentrations if silica is absent ($D_{REG}$, dashed) or present ($D_{SAS}$, solid).

Surface coverage rises, peaking at approximately 2 mg m$^{-2}$. This is close to the coverage expected for a densely packed monolayer, which Koutsopoulos calculated to be 2.1 mg m$^{-2}$. Afterwards, due to trypsin’s autolysis and weak affinity of produced fragments to the surface, surface coverage decays, dropping to almost 0 mg m$^{-2}$.

S5.2 P does not bind to or inhibit trypsin (T)

Northrop mentioned trypsin inhibition by hydrolysis products in 1921.$^5$ Fraser et al. studied trypsin self-digest and suggested that P can form a complex with T, thus inhibiting trypsin, in line with Northrop’s proposal.$^6$ We neglect this inhibition process for three reasons:

1. We lack information such as the binding constant of P to T. Moreover, the degree of fragmentation is unclear, i.e. how many molecules of (potentially inhibiting) P are generated by digestion of one T.

2. Over 24 hours of regular autolysis, trypsin activity drops to almost zero.$^7$

3. Inhibition by P should affect regular and surface accelerated self-digest in the same way, therefore it is of little interest to the project at hand.
S5.3 Degradation of T to P is irreversible

Kukor et al. found trypsin to resynthesize the R117-V118 peptide bond.\textsuperscript{8} Cleavage of R117-V118 appears to be an important step in the degradation of trypsin;\textsuperscript{7} resynthesis of this particular bond is likely to be necessary for reversing inactivation. Neglecting the back reaction is acceptable as experiments show that bulk reaction (BR) trypsin self-digest continues until barely any activity is left.\textsuperscript{7}

S5.4 There is no uncertainty on the x-axes of the plots

Uncertainty in time ($D_{\text{REG}}$ and $D_{\text{SAS}}$) can be neglected because it would be small in relative terms. In $D_{\text{ADS}}$, uncertainty for the x-axis (equilibrium concentration of [T]) should be comparable to the uncertainty in the y-axis in $D_{\text{REG}}$ and $D_{\text{SAS}}$ but we do not include it in our models for simplicity.

S5.5 Trypsin conformations A and B are in equilibrium at $t = 0$ in the kinetic experiments

JW1981’s experimental setup for the kinetic measurements of $D_{\text{SAS}}$ includes a change in pH from 2-3 in the stock solution to 8 at $t = 0$. If their equilibrium constant $K_{\text{Eq}}$ depends on pH, A and B would not be in equilibrium at pH=8. Lacking this information, we assume that $K_{\text{Eq}}^{\text{pH}=2-3} = K_{\text{Eq}}^{\text{pH}=8}$.

S5.6 All trypsin species ($\{A, B, AS, BS\}$ or $\{T, TS\}$) are equally active towards BAEE

We derive experimental trypsin concentrations from relative measurements of trypsin activity towards BAEE (Section S1). The question if A and B (and their adsorbed forms) are equally active towards BAEE is by itself unrelated to the mechanism of (surface accelerated)
autolysis. Nevertheless, it is relevant to ensure comparability of modeled concentrations to experimental data.

Gabel and Kasche described that partially unfolded trypsin states exist and retain their enzymatic activity.\(^9\) The same authors found that Ca\(^{2+}\) leads to a more compact trypsin conformation and reduced autolysis but similar activity towards BAEE.\(^{10}\) Therefore, assuming A and B to be equally active is plausible.

Kunitz and Northrop found that trypsin deactivates reversibly and assumed that this deactivated form is digested by the remaining active trypsin.\(^{11}\) They monitored activity towards hemoglobin instead of BAEE. If the deactivated form (which would correspond to B in our notation) is present in large concentrations, our derivation of experimental trypsin concentrations is wrong. We therefore repeated the inference discussed in the main text, assuming that B is completely inactive towards BAEE and absent at \(t = 0\), i.e. \([T]_0 = [A]_0\). This model also explains the complete heterogeneous data of JW1981 (results not shown).

JW1981 observed no change in activity upon adsorption on silica surfaces. Note that this might not always be the case: results by Koutsopoulos et al. indicate a drop or even complete loss in activity for trypsin adsorbed on silica wafers.\(^2\)

**S6  Bayes inference**

Bayes theorem is

\[
p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)},
\]

where:

\(D\) = experimental data

\(\theta\) = model parameters

\(p(x|y)\) = conditional probability of \(x\) given \(y\)
Bayesian data integration (BDI) of heterogeneous data sets $D = \{D_1, \ldots, D_n\}$ requires individual likelihood functions $p_i$ for every subset $D_i$. The total likelihood is the product of all individual likelihoods

$$p(D|\theta) = \prod_i p_i(D_i|\theta_i).$$

(S33)

$\theta_i$ is a subset of the overall parameter space $\theta$. One can also interpret this approach as Bayesian inference for one data set, using the posterior that we have determined from all other data sets as prior:

$$p(\theta|D_i) = \frac{p_i(D_i|\theta) \prod_{j \neq i} p(\theta|D_j)}{p(D_i)}$$

(S34)

All of our models return concentrations. The likelihood function assumes a normal distribution (Eq. S35). For the current study, the assumption of normality seemed safe, which is supported by the close match of measured data and posterior predictions. If concentrations were close to zero, relative to $\sigma$, using a normal distribution might lead to uncertainty intervals and posterior predictions of negative concentrations which are unphysical.

$$[X]_{t}^{\exp} \sim N([X]_{t}^{\text{model}}, \sigma)$$

(S35)

where:

- $[X]_{t}^{\exp}$ = experimental concentration of X at time t
- $[X]_{t}^{\text{model}}$ = modeled concentration of X at time t
- $\sigma$ = standard deviation

$$[X]_{t}^{\text{model}} = [X]_0 + \int_0^t \frac{d[X]}{dt} dt$$

(S36)

$\frac{d[X]}{dt}$ depends on the kinetic model and its parameters. $[X]^{\text{model}} = [T] + [TS] = [A] + [B] +$
$[AS] + [BS]$ for $D_{REG}$ and $D_{SAS}$ (see also Section S5.6). For $D_{ADS}$, $[X]^{\text{model}} = [TS]$ (there is no time-dependence in this experiment) according to the Langmuir isotherm in Equation 1 (main text).

### S6.1 Influence of $\sigma$

$\sigma$ can be a constant or a nuisance parameter. JW1981 do not provide any error estimate, i.e. $\sigma$ is unknown and treated as a nuisance parameter according to the emcee tutorial. \(^2\)

$$\sigma = \sqrt{\sigma_{\text{fix}}^2 + f^2y_{\text{model}}^2} \quad \text{(S37)}$$

where:

$\sigma_{\text{fix}}$ = underestimated standard deviation

$f$ = fractional deviation

Both $\sigma_{\text{fix}}$ and $f$ are inferred from the data. $\sigma_{\text{fix}}$ represents an absolute error (M) whereas $f$ accommodates relative errors (unitless) dependent of the observed concentration.

If the experimental uncertainty is the same for two experiments $i$ and $j$, $\sigma_i$ should equal $\sigma_j$. For JW1981, we assume that the experimental uncertainty of trypsin concentrations with or without silica surfaces is identical, hence $\sigma_{D_{SAS}} = \sigma_{D_{REG}}$. The experimental uncertainty $\sigma_{D_{ADS}}$ of the adsorption isotherm is unrelated and inferred independently.

### S6.2 Priors

We use weakly informative uniform priors for all parameters of the kinetic models (shown in Figure 2b in the main text). Rate constants of chemical reactions span many orders of magnitude. Parameters are supposed to be restrained by experimental data; priors often

\(^2\text{https://emcee.readthedocs.io/en/latest/tutorials/line/; accessed at 13:48, January 9th, 2020}\)
help numerical inference to converge, or at least they avoid wasting computational time in flat areas of the posterior landscape.

**S6.2.1 Likelihoods are indifferent if step is not rate-determining**

Consider the following simple mechanism:

\[ X \overset{k_1}{\rightarrow} Y \overset{k_2}{\rightarrow} Z \quad (S38) \]

Let’s assume concentrations over time are only available for X and Z, and that the first step is rate-determining. The likelihood is unresponsive to \( k_2 \) as long as \( k_2 \gg k_1 \). Therefore, the posterior landscape would be flat. Sampling in this area would be wasteful and uninformative. A reasonable prior that truncates \( k_2 \) would simplify sampling. Note that the choice of the prior will impact the observed density (Figure S5). Unlike in many other applications of Bayesian statistics, even an infinite amount of data is not guaranteed to dominate over the prior.

![Figure S5: \( p(\theta|D) \) for different uniform priors.](image-url)
S6.2.2 Biochemical motivation for priors

In biological systems, the association rate ($k_a$) is typically between 3 to 7 whereas the dissociation rate ($k_d$) is between -6 to -1. We added two orders of magnitude to the lower and upper boundary as safety margin. Moreover, our system describes the binding of enzyme molecules to binding sites, not to colloidal particles (CPs). Every such particle has $n_{S/CP}$ binding sites on its surface. In order to yield accurate results, parameters should satisfy

$$k_{a,\text{model}}[T][S] = k_{a,\text{real}}[T][CP].$$  \hfill (S39)

Since

$$[S] = [CP] \cdot n_{S/CP}$$  \hfill (S40)

$k_{a,\text{model}}$ will deviate from $k_{a,\text{real}}$

$$k_{a,\text{real}} = k_{a,\text{model}} \cdot n_{S/CP}.$$  \hfill (S41)

Kinetically perfect enzymes have specificity constants $\frac{k_{cat}}{K_m}$ (merged in our model to rate constant of the bulk reaction ($k_{BR}$)) of up to $10^8 - 10^9 \text{ M}^{-1} \text{s}^{-1}$, so we chose $10^9 \text{ M}^{-1} \text{s}^{-1}$ as upper limit for $k_{BR}$. The lower limit can be derived from the integrated rate law of a second-order reaction

$$\frac{1}{[T]} = k_{BR}t + \frac{1}{[T]_0}$$  \hfill (S42)

and the data. Solving Eq. S42 for $k_{BR}$ yields

$$k_{BR} = \left(\frac{[T]_0}{[T]} - 1\right)t^{-1}[T]_0^{-1}.$$  \hfill (S43)

We know from $D_{REC}$ that for $t = 20000 \text{ s}, [T]_0 = 2.2 \mu\text{M}, \frac{[T]}{[T]_0} \ll 98\%$. After rounding we set

---

3https://www.sprpages.nl/kinetics/association; accessed at 15:48, August 2nd, 2019
the lower end of the uniform prior to $k_{BR} > 10^{-0.5} \text{M}^{-1} \text{s}^{-1}$.

The lower limit for $k'_{ER}$ is derived in a similar fashion from the integrated rate law of the ER mechanism (derivation not shown).

$$k'_{ER} = -t^{-1} \left( \log \frac{[T]_0}{[T]} + K_a^{-1} \left( \frac{1}{[T]_0} - \frac{1}{[T]} \right) \right)$$  \hspace{1cm} (S44)

For $[T]_0 = 2.2 \mu\text{M}$, $[T] < 0.9 \mu\text{M}$ after $t = 15000$ s. Larger $K_a$ lead to smaller $k'_{ER}$, hence we apply the upper limit of $K_a$, $10^{12} \text{M}^{-1}$, yielding a lower limit of $10^{-5.15} \text{s}^{-1}$ for $k'_{ER}$. We define priors for $k'_{ER}$ instead of $k_{ER} = \frac{k'_{ER}}{[S]_0}$ (see Section S2.2) because Equation S44 is independent of $[S]_0$. The upper limit for $k'_{ER}$ cannot be determined without knowing $\phi$ beforehand. We set the upper limit to $10^{-0.15} \text{s}^{-1}$, i.e. the prior spans five orders of magnitude.

For the models including two conformations of trypsin we add two uniform priors:

$$k_{AB} \sim \text{unif}(10^{-8} \text{s}^{-1}, 10^{-2} \text{s}^{-1})$$  \hspace{1cm} (S45)

$$K_{Eq} = \frac{k_{AB}}{k_{BA}} \sim \text{unif}(10^{-8}, 10^0)$$

Priors for nuisance parameters are shown in Table S1.

Table S1: Priors for nuisance parameters

| parameter | lower limit | upper limit | unit |
|-----------|-------------|-------------|------|
| $\sigma_{\text{kin}}$ | $([T] + [TS])_{\text{exp min}} \cdot 10^{-3}$ | $([T] + [TS])_{\text{exp max}}$ | M |
| $f_{\text{kin}}$ | $10^{-4}$ | 1 | |
| $\sigma_{\text{ads}}$ | $[TS]_{\text{min}} \cdot 10^{-3}$ | $[TS]_{\text{exp max}}$ | M |
| $f_{\text{ads}}$ | $10^{-4}$ | 1 | |
S7 Proof of concept: Bayesian inference of parameters for kinetic models

Although there are examples of BDI for systems of ordinary differential equations (ODEs) in related fields such as metabolic networks,\textsuperscript{13} toxicokinetics\textsuperscript{14} and system biology,\textsuperscript{15} we wanted to prove that Bayesian inference can express uncertainty for parameters in chemical kinetics and synergize heterogeneous data sources. To that end, we generated synthetic data with kinetic and thermodynamic models (Figure S6) using subsets of the true parameters ($\theta_{\text{true}}$) for the reactions shown in Scheme S3, aiming to reobtain $\theta$ via Bayesian inference. Data on kinetics ($D_{\text{KIN}}$) display the concentration of a protease A. Equilibrium adsorption ($D_{\text{EQ}}$) data track the adsorbed amount of A on a surface compared to the remaining unbound concentration. Time-dependent adsorption ($D_{\text{TD}}$) data monitor the same adsorption over time before reaching the equilibrium. The assumptions are similar to those discussed for the main paper (Section S5). Jupyter notebooks containing the models, data generation and the Bayesian parameter inference can be found on GitHub, together with all the required code: https://github.com/niklastoe/kineticmodel_bdi.

Scheme S3: All reactions modeled in the proof of concept. The reaction in green is only possible in $D_{\text{KIN}}$. In both $D_{\text{EQ}}$ and $D_{\text{TD}}$, only adsorption and desorption are possible. This is equivalent of blocking the green reaction with an inhibitor.

Considering multiple data sets during sampling of the parameter space leads to a drastic reduction of uncertainty (Figure S7). If $D = \{D_{\text{KIN}}\}$, parameter uncertainty covers multiple orders of magnitude. In most cases, only the priors limit the range, proving once more that completely different sets of parameters can yield similar results. Including $D_{\text{EQ}}$ in BDI tightly restrains $K_a$ and $[S]_0$. $D_{\text{TD}}$ restrain $[S]_0$, the association rate ($k_a$) and to some extent
Figure S6: Three synthetic data sets based on $\theta_{\text{true}}$. Lines correspond to the noise-free values generated from the models based on $\theta_{\text{true}}$, markers include added noise. In $D_{TD}$, two starting concentrations of A ([A]$_0$) were studied.

the dissociation rate ($k_d$). Combining all three data sets, the accessible parameter space shrinks and $k_{ER}' = k_{ER}[S]_0$ (see Section S6.2.2) is approximated well.

BDI does not only correctly infer kinetic parameters but can also distinguish between different mechanisms (Figure S8). In this test case, we compare two models $M$: ER and LH. Both mechanisms can reproduce $D_{KIN}$ equally well which is not surprising. Due to the added noise, $p(\theta|D)$ can be larger than $p(\theta_{\text{true}}|D)$. This is a sign of (weak) overfitting: since the noise is random, no set of parameters can be consistently better than $\theta_{\text{true}}$. If we include $D_{EQ}$, LH still performs similar to ER. Finally, with the inclusion of the last data set $D_{TD}$, the difference in $p(\theta|D,M)$ between ER and LH becomes large. The LH mechanism cannot reproduce all available data as well as the ER mechanism and is therefore rejected for our test case.

Our example for synthetic data demonstrates that integrative Bayesian inference reduces parameter uncertainty. Leveraging multiple data sets allows to reject a wrong hypothetical
Figure S7: Marginalized posterior distributions $p(\theta|D)$ for $D = \{D_{\text{KIN}}\}$ (green), $D = \{D_{\text{KIN}}, D_{\text{EQ}}\}$ (orange), $D = \{D_{\text{KIN}}, D_{\text{EQ}}, D_{\text{TD}}\}$ (blue). True underlying parameter values $\theta_{\text{true}}$ are shown as black lines, priors are shown as red dashed lines. Refer to Section S6.2.2 for details why the prior is defined for $k'_{\text{ER}} = k_{\text{ER}}[S]_0$.

Figure S8: The maximum log posterior for ER and LH depends on the inclusion of different data sets.
mechanism. In our example, the LH mechanism could only be rejected after uncertainty was reduced for all parameters involved in the mechanism. Of course, this may pose a challenge in practice, depending on the accessibility and cost of experiments for parameter inference.

References

(1) Rohatgi, A. WebPlotDigitizer. 2019; https://automeris.io/WebPlotDigitizer.

(2) Koutsopoulos, S.; Patzschly, K.; Bosker, W. T.; Norde, W. Adsorption of trypsin on hydrophilic and hydrophobic surfaces. *Langmuir* **2007**, *23*, 2000–2006.

(3) Cecilia Hahn Berg, I.; Muller, D.; Arnebrant, T.; Malmsten, M. Ellipsometry and TIRF studies of enzymatic degradation of interfacial proteinaceous layers. *Langmuir* **2001**, *17*, 1641–1652.

(4) Nord, F. F.; Bier, M.; Terminiello, L. On the mechanism of enzyme action. LXI. The self digestion of trypsin, calcium-trypsin and acetyltrypsin. *Arch. Biochem. Biophys.* **1956**, *65*, 120–131.

(5) Northrop, J. H. The Inactivation of Trypsin: II. The Equilibrium between Trypsin and the Inhibiting Substance . Influence of the Order of Mixing and of the Time of Standing on the. *J. Gen. Physiol.* **1922**, *227*, 245–260.

(6) Powell, E.; Fraser, D. The Kinetics of Trypsin Digestion. *J. Biol. Chem.* **1950**, *187*, 803–820.

(7) Várallyay, É.; Pál, G.; Patthy, A.; Szilágyi, L.; Gráf, L. Two mutations in rat trypsin confer resistance against autolysis. *Biochem. Biophys. Res. Commun.* **1998**, *243*, 56–60.

(8) Kukor, Z.; Tóth, M.; Pál, G.; Sahin-Tóth, M. Human Cationic Trypsinogen. *J. Biol. Chem.* **2002**, *277*, 6111–6117.
(9) Gabel, D.; Kasche, V. Cooperative transitions between active $\alpha$- and $\beta$-trypsin conformations. Biochem. Biophys. Res. Commun. 1972, 48, 1011–1018.

(10) Gabel, D.; Kasche, V.; Soininen, R.; Waisvisz, J. M.; van der Hoeven, M. G.; Swahn, C.-G. Autolysis of beta-Trypsin. Influence of Calcium Ions and Heat. Acta Chem. Scand. 1973, 27, 1971–1981.

(11) Kunitz, M.; Northrop, J. H. Inactivation of Crystalline Trypsin. J. Gen. Physiol. 1934, 17, 591–615.

(12) Berg, J. M.; Tymoczko, J. L. L. Stryer, 2002. Section 8.4: The Michaelis-Menten Model Accounts for the Kinetic Properties of Many Enzymes. Biochem. 5th Ed. Berg, JM, Tymoczko, JL L. Stryer (Eds.), WH Free. New York, ISBN-10 0-7167-3051-0

(13) Liebermeister, W.; Klipp, E. Bringing metabolic networks to life: integration of kinetic, metabolic, and proteomic data. Theor. Biol. Med. Model. 2006, 3, 42.

(14) Ratier, A.; Lopes, C.; Labadie, P.; Budzinski, H.; Delorme, N.; Quéau, H.; Peluhet, L.; Geffard, O.; Babut, M. A Bayesian framework for estimating parameters of a generic toxicokinetic model for the bioaccumulation of organic chemicals by benthic invertebrates: Proof of concept with PCB153 and two freshwater species. Ecotoxicol. Environ. Saf. 2019, 180, 33–42.

(15) Thijssen, B.; Dijkstra, T. M.; Heskes, T.; Wessels, L. F. Bayesian data integration for quantifying the contribution of diverse measurements to parameter estimates. Bioinformatics 2018, 34, 803–811.