Insights Into Accelerated Protease Autolysis and Classifier Uncertainty

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

Dr. rer. nat.

der Fakultät für

Biologie

an der

Universität Duisburg-Essen

vorgelegt von

Niklas Tötsch

aus Marl

Oktober 2020

DOI: 10.17185/duepublico/74172
URN: urn:nbn:de:hbz:464-20210420-093510-2

Die der vorliegenden Arbeit zugrunde liegenden Experimente wurden in den Abteilungen für Bioinformatik des Zentrums für Medizinische Biotechnologie (ZMB) der Universität Duisburg-Essen durchgeführt.

1. Gutachter: Prof. Dr. Daniel Hoffmann
2. Gutachter: Prof. Dr. Bettina Siebers
3. Gutachter: Prof. Dr. Katja Ickstadt

Vorsitzender des Prüfungsausschusses: Prof. Dr. Peter Bayer
Tag der mündlichen Prüfung: 19.02.2021

Contents

Summary

Proteases cleave peptide bonds in proteins and play a crucial role in various physiological functions such as digestion and post-translational modification. Poor regulation can cause medical emergencies by e.g. enzymatic digestion of gastrointestinal walls. Conventional small-molecule protease inhibitors usually rely on blocking the protease's active site. A radically different approach exploits the enyzme's autolysis. The presence of certain polymers or colloids accelerates the cannibalistic digest, which effectively eliminates the proteolytic activity in a short period of time. Developing inhibitors that take advantage of this phenomenon may be a promising alternative to conventional inhibitors for medical or biotechnological use. A detailed understanding of the acceleration is therefore necessary.

This thesis starts with an introduction to proteases, their conventional inhibitors and the novel inhibition strategy (section 1.1). Since this thesis heavily features Bayesian statistics, its central ideas and advantages are presented in chapter 2. The main results of my doctoral work are summarized in my publications (chapter 3).

The first study (section 3.1) is a joint experimental/computational investigation of accelerated self-digest. It extends previous work by Gilles et al., which among other things described substoichiometric inhibition of proteases through polymers. Our new study proves that polymers can accelerate trypsin's self-digest under physiological conditions. Computational surface scanning indicates that the polymers bind multivalently to trypsin's surface without blocking the active site or all cleavage sites. We hypothesize that the polymers serve as template where many trypsin molecules bind, leading to a drastically increased local concentration that would explain the accelerated self-digest. This local enrichment is also modeled computationally.

While this hypothesis is plausible and in agreement with the experimental and computational findings, more evidence is needed to prove it. Johnson and Whateley suggested a different explanation for the accelerated loss of trypsin activity they observed when colloidal silica or negatively charged polymers were present. They assumed that adsorbed trypsins are more susceptible to autolysis. According to their hypothesis, the substrate trypsins are adsorbed while the digester trypsins attack from solution. The reaction cannot be monitored experimentally on a molecular level. Simulations are infeasible since the system, including at least two trypsin molecules and a polymer or colloid, is too large for combined quantum mechanical/molecular mechanical techniques, which would be required to observe the cleavage of the substrate. Therefore, kinetic modeling is the most viable approach to compare the rivaling hypotheses. Since kinetic models are often under-determined, I developed an approach for Bayesian data integration to include all available kinetic and thermodynamic data from Johnson's publication (section 3.2). While this study falsifies several of Johnson's assumptions and conclusions, the available data is insufficient to disprove the increased susceptibility hypothesis or the local enrichment hypothesis.

In addition to the work on accelerated self-digest, this thesis is also concerned with the evaluation of classifiers. Prediction if a molecule is a protease inhibitor or a non-inhibitor

would be an example of classification. The basic concepts of classification are explained in section 1.2. Small sample sizes make it difficult to definitively measure a classifier's performance. In many classification experiments, this is ignored. I developed a Bayesian approach to determine the uncertainty of all classification performance metrics that can be calculated from the confusion matrix (section 3.3). In addition, this work shows how large the uncertainty in published classifiers can be due to small data sets, which are common in biology or chemistry. Some of them could be worse than random guessing. To facilitate the design of new classification experiments, a simple rule that allows to estimate the necessary sample size beforehand is derived.

Zusammenfassung

Proteasen spalten Peptidbindungen in Proteinen und spielen eine entscheidende Rolle bei verschiedenen physiologischen Funktionen wie der Verdauung und der posttranslationalen Modifikation. Eine schlechte Regulierung kann medizinische Notfälle verursachen, z.B. durch enzymatische Verdauung der Wände des Magen-Darm-Trakts. Herkömmliche Proteaseinhibitoren beruhen in der Regel auf der Blockierung des aktiven Zentrums der Protease. Ein völlig anderer Ansatz nutzt die Autolyse des Enzyms aus. Die Präsenz bestimmter Polymere oder Kolloide beschleunigt den kannibalistischen Verdau, wodurch die proteolytische Aktivität in kurzer Zeit wirksam eliminiert wird. Die Entwicklung von Inhibitoren, die sich dieses Phänomen zunutze machen, kann eine vielversprechende Alternative zu herkömmlichen Inhibitoren für den medizinischen oder biotechnologischen Gebrauch sein. Ein detailliertes Verständnis der Beschleunigung ist daher notwendig.

Diese Arbeit beginnt mit einer Einführung über Proteasen, ihre konventionellen Inhibitoren und die neuartige Inhibitionsstrategie (Abschnitt 1.1). Da diese Arbeit stark von der Bayes'schen Statistik geprägt ist, werden ihre zentralen Ideen und Vorteile in Kapitel 2 dargestellt. Die wichtigsten Ergebnisse meiner Doktorarbeit sind in meinen Publikationen (Kapitel 3) zusammengefasst.

Die erste Studie (Abschnitt 3.1) ist eine gemeinsame experimentelle/computergestützte Untersuchung der beschleunigten Autolyse. Sie erweitert frühere Arbeiten von Gilles et al., die u. a. die substöchiometrische Hemmung von Proteasen durch Polymere beschrieben. Unsere neue Studie beweist, dass Polymere den Selbstverdau von Trypsin unter physiologischen Bedingungen beschleunigen können. Das computergestützte Oberflächenscreening zeigt, dass die Polymere multivalent an die Trypsinoberfläche binden, ohne das aktive Zentrum oder alle Spaltstellen zu blockieren. Wir stellen die Hypothese auf, dass die Polymere als Templat dienen, auf dem viele Trypsinmoleküle binden, was zu einer drastisch erhöhten lokalen Konzentration führt. Somit würde der beschleunigte Selbstverdau erklärt. Diese lokale Anreicherung wird zusätzlich mit einem Computermodell untersucht.

Diese Hypothese ist zwar plausibel und stimmt mit den experimentellen und simulierten Ergebnissen überein, doch sind weitere Beweise erforderlich, um sie zu belegen. Johnson und Whateley schlugen eine andere Erklärung für den beschleunigten Verlust der Trypsinaktivität vor, den sie beobachteten, wenn kolloidales Siliciumdioxid oder negativ geladene Polymere vorhanden waren. Sie nahmen an, dass auf dem Kolloid adsorbierte Trypsine anfälliger für Autolyse sind. Ihrer Hypothese zufolge sind die adsorbierten Trypsine das Substrat, während die verdauenden Trypsine aus der Lösung angreifen. Die Reaktion lässt sich experimentell nicht auf molekularer Ebene verfolgen. Simulationen sind nicht durchführbar, da das System, das mindestens zwei Trypsinmoleküle und ein Polymer oder Kolloid enthält, für kombinierte quantenmechanische/molekülmechanische Simulationen, die zur Beobachtung der Spaltung des Substrats erforderlich wären, zu groß ist. Daher ist die kinetische Modellierung der naheliegendste Ansatz, um die beiden Hypothesen zu vergleichen. Da kinetische Modelle oft unterbestimmt sind, habe ich einen Ansatz für Bayes'sche Datenintegration entwickelt, um alle verfügbaren kinetischen und thermodynamischen Daten aus Johnsons Publikation einzubeziehen (Abschnitt 3.2). Während

diese Studie mehrere Annahmen und Schlussfolgerungen von Johnson widerlegt, reichen die verfügbaren Daten nicht aus, um eine der beiden Hypothesen (lokale Anreicherung oder erhöhten Anfälligkeit für Verdau) auszuschließen.

Neben den Arbeiten zur beschleunigten Autolyse beschäftigt sich diese Arbeit auch mit der Bewertung von Klassifikatoren. Die Vorhersage, ob ein Molekül ein Proteaseinhibitor ist oder nicht wäre ein Beispiel für eine Klassifizierung. Die grundlegenden Konzepte der Klassifikation werden in Abschnitt 1.2 erläutert. Kleine Stichprobengrößen machen es schwierig, die Leistung eines Klassifikators genau zu messen. In vielen Klassifikationsexperimenten wird dies ignoriert. Ich habe einen Bayes'schen Ansatz entwickelt, um die Unsicherheit aller Leistungskennzahlen der Klassifikation zu bestimmen, die sich aus der Konfusionsmatrix berechnen lassen (Abschnitt 3.3). Darüber hinaus zeigt diese Arbeit, wie groß die Unsicherheit in veröffentlichten Klassifikatoren aufgrund kleiner Datensätze sein kann, die in der Biologie oder Chemie häufig vorkommen. Einige der Klassifikatoren könnten schlechter sein als einfaches Raten. Um das Design neuer Klassifikationsexperimente zu erleichtern wird eine einfache Regel hergeleitet, die es erlaubt, die erforderliche Stichprobengröße im Voraus zu schätzen.

Chapter 1

Introduction

1.1 Accelerated Protease Self-digest as Inhibition Strategy

1.1.1 Proteases and Conventional Inhibition

Proteases are enzymes that catalyze the cleavage of peptide bonds in proteins (Figure 1.1). This step may lead to complete degradation of the substrate, for example during digestion. In other scenarios, notably proteolytic post-translational modification, it is a limited and highly directed hydrolysis of a peptide bond to alter the protein's structure, enabling new or modified biological activity. [1]

The catalytic activity of a protease is governed by the active site and its surrounding. Substrate-binding pockets allow proteases to cleave selectively after certain patterns in the substrate's amino acid sequence. For example, trypsin is known to cleave peptide bonds predominantly C-terminally to lysine and arginine. This is caused by the presence of an aspartate residue (D189) in trypsin's selectivity pocket, which favors interactions with positively charged residues of the substrate. In chymotrypsin, this selectivity pocket is more hydrophobic. Therefore, chymotrypsin cleaves C-terminally to hydrophobic residues such as tyrosine, phenylalanine and tryptophan.

Due to their effect on proteins in their environment, poor regulation of proteases can have dramatic consequences. In patients suffering from pancreatitis, proteases such as trypsin digest the structural cells of the pancreas. [2] Complications from pancreatitis are rare. Nevertheless, they can be severe and potentially lethal. [3] Moreover, repeated inflammation of the pancreas is suspected to promote the development of pancreatitic

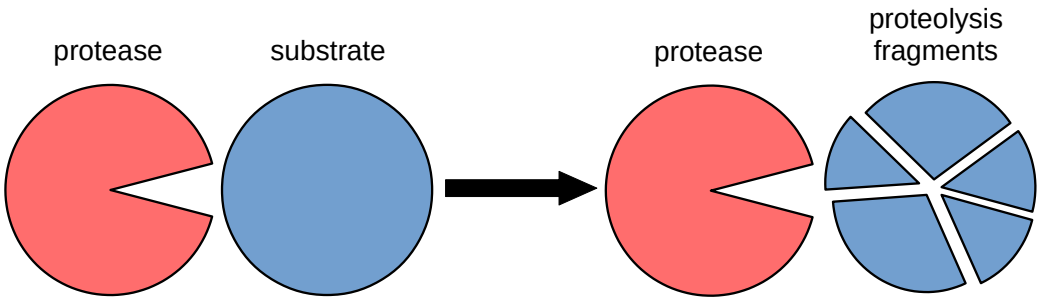


Figure 1.1: Schematic of proteolysis. During proteolysis, a protease digests another protein. The substrate is broken into smaller pieces. If the substrate is also a molecule of the same protease, this reaction is called autolysis.

cancer. [2] Effective and targeted inhibitors could play a pivotal role in therapies regulating proteolytic activity. In addition, protease inhibitors might also have economic value. Industrial applications of proteases include protein stain removal, milk clotting for dairy production, or biofilm removal in the pulp and paper industry. [4] Inhibitors allow to control and eventually terminate the enzymatic reaction at the desired time point.

In spite of the structural differences causing the proteolytic selectivity mentioned above, the active site is similar in proteases of the same type. For example, serine proteases, a family of proteases, all share the same catalytic triad (consisting of serine, histidine, and aspartic acid), which breaks the peptide bond of the substrate. Consequently, classic protease inhibitors often inhibit several related proteases. Aprotinin, a well known protease inhibitor administered as a drug to reduce bleeding during complex surgery (trade name Trasylol), inhibits among others trypsin, chymotrypsin and kallikrein. [5] While inhibition of multiple proteases might be desirable in some applications, in others one would prefer to selectively target those enzymes that need to be downregulated without interfering with the other ones. Another drawback of conventional protease inhibitors lies in the need for stoichiometric concentrations. One inhibitor molecule can only inhibit a single enzyme. Often, an excess inhibitor concentration is needed to ensure the necessary decrease in protease activity. Larger concentrations increase the likelihood of unwanted side effects..

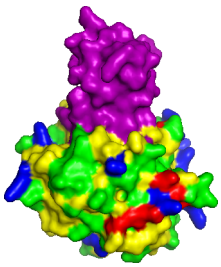
1.1.2 Autolysis: From Nuisance to Inhibition Strategy

In order to overcome those two limitations, a completely different inhibition strategy has been proposed. Autolysis, cannibalistic self-digest of proteases, is often considered an experimental nuisance because the protease concentration decreases over time. [6, 7] Interestingly, the presence of certain polymers or colloids accelerates self-digest of trypsin, chymotrypsin and savinase. [7, 8, 9, 10] Lv et al. proposed to exploit this phenomenon to control protease activity. [10] Strongly substoichimetric concentrations of their polymers were sufficient to eliminate proteolytic activity almost completely within minutes.

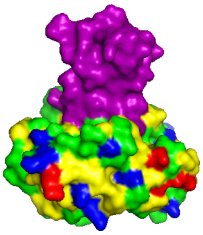
Unfortunately, these polymers are not necessarily specific for targeted proteases. Lv et al. state that novel polymers are needed that bind to selected enzymes exclusively in order to develop radically new therapies for poorly regulated proteases. [10] Accelerated self-digest (AS) relies on the binding towards the protein surface. The active site must remain unblocked so that proteases can digest each other. Contrary to the well conserved active sites, the protease surfaces are much more diverse (Figure 1.2). Multivalent ligands, such as polymers with affinity side-chains, can form cooperative interactions through the combination of diverse non-covalent bonds. [11] Tailored ligands matching the unique pattern on the targeted protease surface are therefore promising candidates for selective inhibition.

Polymers can be synthesized from a combinatorial library of monomers. This allows for a tremendous chemical space. Gilles presented such polymers; several of them inhibit proteases. [17] Some of these polymers were selective towards one or two proteases out of a total of six. These findings are encouraging and confirm the potential of this approach.

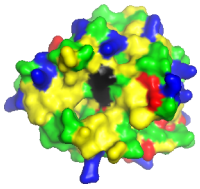
The design of novel selective inhibitors can be assisted by modeling. Physics based simulations show how these polymers bind but they consume lots of computational time, which limits the number of candidate polymers that can be tested. [18] Statistical modeling does not provide molecular details but it is orders of magnitude faster. I built a Bayesian logistic regression model that predicts the inhibition for all six proteases from Gilles original data (unpublished, results not shown). Only the composition of the polymers, that is the share of the different monomers from the combinatorial library, is considered as



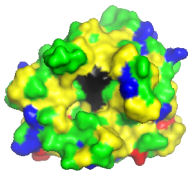
(a) Chymotrypsin side view



(b) Trypsin side view



(c) Chymotrypsin top view



(d) Trypsin top view

Figure 1.2: Protease surface renderings [12] of chymotrypsin (PDB 1CBW [13]) and trypsin (PDB 2FTL [14]) at pH 8.0. [15, 16] Positive residues are shown in blue, negative residues in red, hydrophilic residues in green and hydrophobic residues in yellow. The catalytic triad (H57, D102, S195) is shown in black. Aprotinin is shown in purple in the upper subfigures but hidden in the lower ones.

input. Leave-one-out cross-validation shows that the model can reproduce experimental data surprisingly well. Consequently, the inhibitory potential of completely new polymer compositions can be tested with this approach.

Since the model predicts if a polymer inhibits any of the proteases, it can also predict if a polymer would be a selective inhibitor. This task is substantially more difficult than the correct prediction of the inhibition for a single protease because the prediction could be wrong for any of the enzymes. Further experiments would be needed to prove that my model can reliably predict selective inhibitors. If it does, the model could speed up the search for selective protease inhibitors tremendously.

The statistical model does not consider how the polymers inhibit the proteases. Most of Gilles' inhibiting polymers presumably block the active site, either by binding there directly or by sterically restricting access, similarly to conventional inhibitors. Yet, others inhibit proteases at strongly substoichiometric concentrations, suggesting that they might accelerate autolysis rather than block multiple protease molecules simultaneously. Combining the concepts of autolysis acceleration with the selectivity of multivalent ligands might therefore overcome both limitations of conventional protease inhibitors.

In the first publication (section 3.1), joint experimental and computational work proves that two polymers consisting of amino acid-selective binding monomers accelerate the autolysis of trypsin, paving the way towards a new class of selective protease inhibitors. One obstacle during the design of such autolysis accelerators is the limited mechanistic understanding of the phenomenon. Johnson and Whateley hypothesized that adsorbed proteases are more susceptible to proteolysis. [8] In the second project on AS (section 3.2), we investigated the reaction mechanism with a combination of kinetic modeling and Bayesian data integration. In this way, we could test mechanistic hypotheses against all available data, which was not the case in Johnson and Whateley's original publication.

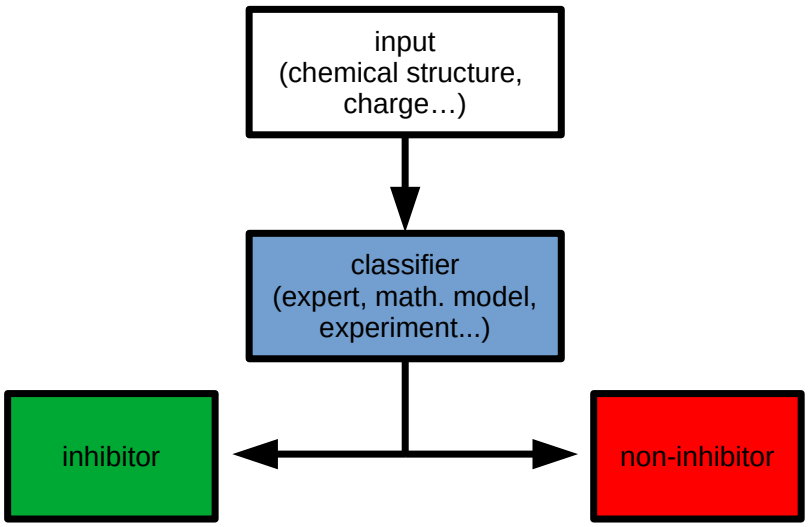


Figure 1.3: Schematic of classification. A classifier assigns class labels based on the given input.

1.2 Classification

In statistics, assigning a class label to a data point is called classification. For example, the prediction if a molecule inhibits a protease can be considered a classification. The molecule is labeled either as an *inhibitor* or *non-inhibitor*, the two possible classes. Classification requires input, a classifier and a definition of the classes. The input is passed to the classifier (Figure 1.3). In principle, anything that maps the input to a label could be a classifier. For example, a scientist can assign labels based on the molecule's chemical structure, their experience and chemical intuition. Often, a classifier is a mathematical or statistical model that outputs either a discrete label or a continuous score for these labels. The scores can be converted to a label using thresholds. For the sake of simplicity, this thesis focuses on binary classification, which means that there are only two classes. All concepts discussed in this work can be generalized to multiclass classification.

The benefits of classifiers are clear in our example: the number of molecules that could be protease inhibitors is much larger than the experimental capacities to synthesize and test them. Classifiers can help with the preselection of promising candidates, saving time and resources. However, the risks of using a classifier, for example the consequences of a protease inhibitor being wrongfully discarded, can be unclear. Yet, they could potentially be high. The development of novel protease inhibitors could become more costly or require more time, which would have negative medical and economic effects as described in section 1.1. In other high-stakes environments, poor classifier performance has already caused immediate and major harm to individuals and society:

- A classifier trained by the Michigan Unemployment Insurance Agency to identify fraud completely failed. It lead to false charges in tens of thousands of cases, and massive damages for the wrongly accused.[1]

- Bite mark analysis has been used in court in the US since at least 1870. [19] Among other things, it is used to determine if a bite mark on a victim of a crime was

[1] `https://www.theatlantic.com/technol-ogy/archive/2020/06/michigan-unemployment-fraud-automation/612721/`; accessed on August 27th, 2020, 19:05 CEST

		Reality / class		
		inhibitor (+)	non-inhibitor (-)	Total
Prediction / label	inhibitor (+)	TP = 9	FP = 2	11
	non-inhibitor (-)	FN = 1	TN = 8	9
	Total	10	10	$N = 20$

(a) Example of the results of a classification experiment.

Metric	Abbr.	Alias	Definition	Value
prevalence	-	-	$\frac{\text{TP+FN}}{N} = p(class = +)$	50%
accuracy	ACC	-	$\frac{\text{TP+TN}}{N} = p(class = label)$	85%
true positive rate	TPR	sensitivity	$\frac{\text{TP}}{\text{TP+FN}} = p(label = +\|class = +)$	90%
true negative rate	TNR	specificity	$\frac{\text{TN}}{\text{TN+FP}} = p(label = -\|class = -)$	80%
positive predictive value	PPV	precision	$\frac{\text{TP}}{\text{TP+FP}} = p(class = +\|label = +)$	82%
negative predictive value	NPV	-	$\frac{\text{TN}}{\text{TN+FN}} = p(class = -\|label = -)$	89%

(b) Classification performance is evaluated with metrics that are calculated from the confusion matrix.

Table 1.1: Results and evaluation of an exemplary classification experiment.

caused by the suspect. This can be formulated as a classification problem where the labels are "bite mark was caused by the suspect" and "bite mark was not caused by the suspect". However, it is disputed if this classification is reliable. In 2009, a report by the United States National Academy of Science found "no evidence of an existing scientific basis for identifying an individual to the exclusion of all others" by bite mark analysis. [20] In the United States, at least two dozen convicts, many imprisoned for murder, were exonerated after DNA tests proved them innocent.[2] Their verdicts were at least partially based on bite mark analysis. Several of them had spent decades in prison.

1.2.1 Performance Metrics

Due to these severe consequences, a detailed understanding of a classifier's performance and likelihood of misclassification is essential. The results of a classification experiment can be visualized in a confusion matrix (CM) (Table 1.1a). This gives an overview of true positive (TP), false positive (FP), false negative (FN) and true negative (TN) classifications.

Since the individual entries of the CM do not have a meaningful interpretation with respect to the classifier properties, the performance of a classifier is usually evaluated based on metrics (Table 1.1b). Perhaps the most common metric is accuracy. It measures the share of correctly classified instances out of all data. Accuracy is easy to understand

[2] `https://www.innocenceproject.org/wp-content/uploads/2019/01/Description-of-bite-mark-exonerations-and-statistical-analysis_UPDATED-01.28.19.pdf`; accessed on August 27th, 2020, 19:35 CEST

but it can be misleading. If the data set is imbalanced, which means that the prevalence of the positive class is either higher or lower than 50%, a high accuracy does not guarantee that the classifier is good. Class imbalance is common: testing molecules for inhibition one would expect that there are much more candidates that have no effect on the protease than inhibitors. A "stupid" classifier that predicts *non-inhibitor* in all cases would have a high accuracy, yet its practical value would be zero because its predictions are not informed.

Therefore, accuracy should not be used without second thought. The correct choice of metric largely depends on the data set and the goal of the classification. The most basic metrics determine the share of correct labels out of a subset of the confusion matrix. As can be seen in Table 1.1a, a molecule has a real class (columns) and a label assigned by the classifier (rows). The true positive rate (TPR) measures how many molecules that belong to the *class* inhibitor are correctly labeled (Table 1.1b). Analogously, the true negative rate (TNR) represents how many real non-inhibitors are truthfully detected. TPR and TNR fully describe a classifier's behavior because they tell us the likelihoods for the correct labels based on the true class of the input. They are not biased by class imbalance. [21] On the other hand, the positive predictive value (PPV) quantifies the share of inhibitor *labels* that are correct. The negative predictive value (NPV) does the same for the share of non-inhibitor labels. PPV and NPV describe how reliable the predicted labels are which depends both on the quality of the classifier as well as on the prevalence of the classes in the data set. Consequently, one should keep in mind that PPV and NPV are biased by class imbalance, similarly to accuracy. [21] Finally, it is noteworthy that PPV and NPV can be defined in terms of TPR, TNR, and prevalence. [22]

Ideally, one can decide which of TPR, TNR, PPV, and NPV is the most important for a given task at hand. Let us consider once more the example of predicting novel protease inhibitors from a vast chemical space. Predicting new labels with the classifier is cheap, synthesizing and testing those molecules in the laboratory is expensive. Priorities for the four metrics could look like this:

- PPV (most important). Positive predictions should be correct as much as possible. These molecules will eventually be synthesized. Low precision would lead to a lot of synthesized molecules that turn out to be non-inhibitors, wasting time and money.
- TPR (important). A high TPR means that less actual inhibitors are discarded because the classifier labels them as non-inhibitors. If TPR is too low, the classifier would generate only few *inhibitor* labels. This would reduce the number of promising leads selected for synthesis.
- NPV (unimportant). Molecules with the negative label *non-inhibitor* are not considered further. Therefore, if these labels are correct is of no further importance as long as enough positive labels are generated.
- TNR (unimportant). A high PPV ensures that the classifier does not produce too many false positives relative to the number of true positives, so TNR can be expected to be decent for these purposes.

In this example, it would be reasonable to optimize the classifier primarily for PPV while not letting TPR decrease too much. Since NPV and TNR are not critical for the task at hand, they do not have to be considered. This approach would lead to the best classifier for the assignment. Yet, there are two problems. First, one needs to know one's preferences and priorities well. These might be less obvious for other classification tasks. Second, the obtained classifier is tailored to the task at hand. It would hardly be adequate

if the goal or data set changes. For example, if a subsequent study requires molecules to be tested to verify that they do not hamper the protease's activity, NPV and TNR will be the most important metrics. A classifier that was optimized for high PPV and TPR would be inadequate for this task.

1.2.2 Small Sample Sizes Cause Uncertainty

Another problem in the evaluation of classification experiments lies in the sample size (N), which can be small. For example, the values for the performance metrics given in Table 1.1b are subject to statistical noise. One would like to determine how uncertain these values are to properly understand a classifiers performance. This is also necessary to know if these values can be expected to be reproducible.

There are two publications on Bayesian methods to determine the uncertainty of performance metrics caused by small N. Brodersen's approach only works for some of the metrics. [23] Caelen's method allows to calculate the posterior for all metrics based on the confusion matrix, [24] yet it is only approximately correct for large N and wrong if N approaches zero. In section 3.3, we developed a new approach that is universally applicable to all metrics and correct for all sample sizes.

Furthermore, it is important to know how many predictions must be correct in order to consider a classifier better than random. Ronald Fisher developed his famous exact test to verify if a lady could identify, as opposed to guessing by chance, how her tea was prepared in 1922. [25] While the problem is old, it continues to be overlooked in many studies in the domain of biology and chemistry. Like Fisher's exact test, most methods to determine the impact of small N are frequentist. [26] They are based on null-hypothesis significance testing (NHST), which has considerable limitations (subsection 2.1.4). Bayesian methods may therefore be promising alternatives (subsection 2.1.4).

Chapter 2

Methods

2.1 Bayesian Inference

Bayes' rule allows to update our beliefs with new evidence. It expresses these beliefs in the form of conditional probabilities $p(x|y)$, i.e. the probability of x given y. In Bayesian inference, parameters (θ) of a model are inferred given the observed data (D).

$$\underbrace{p(\theta|D)}_{posterior} = \underbrace{p(\theta)}_{prior} \cdot \underbrace{p(D|\theta)}_{likelihood} / \underbrace{p(D)}_{evidence} \tag{2.1}$$

The posterior $p(\theta|D)$ is the probability distribution of θ given the observations D. Bayesian inference will determine which θ are likely, i.e. have a high posterior probability, based on two types of information: 1) our previous knowledge about the parameters, called the prior $p(\theta)$, and 2) the ability to reproduce the observed D, called the likelihood $p(D|\theta)$. The posterior distribution is a compromise between what we already know about θ and what parameters would fit the observed D well. The evidence is the probability distribution of D. It is defined as

$$p(D) = \int p(\theta) \cdot p(D|\theta) d\theta \tag{2.2}$$

and often difficult to determine. Usually, the evidence does not have to be calculated since the exact posterior is not needed. Rather, one is interested in ratios between $p(\theta'|D)$ and $p(\theta''|D)$ to determine if θ' or θ" is more likely. For this purpose, $p(D)$ can be considered a constant for the observed D and Equation 2.1 simplifies to

$$p(\theta|D) \propto p(\theta) \cdot p(D|\theta). \tag{2.3}$$

Bayesian inference has two major advantages over parameter fitting, for example by maximum likelihood methods. First, the inclusion of prior knowledge can make model parameters more robust because they take additional information into account. Second, parameter fitting only yields a single value for θ whereas Bayesian inference leads to probability distributions. These probability distributions are much more representative. If there are multiple sets of parameters that reproduce the data well, they are captured in the probability distribution. This is especially important if the available data is limited. If there are only few observations, many different parameters can describe the data relatively well. In Bayesian inference, the posterior distribution would be broad. In many cases, the posterior would resemble the prior because little additional information has been taken into account. If more observations are gathered, less parameters match. The posterior distribution narrows. θ can be determined more precisely. With Bayesian inference,

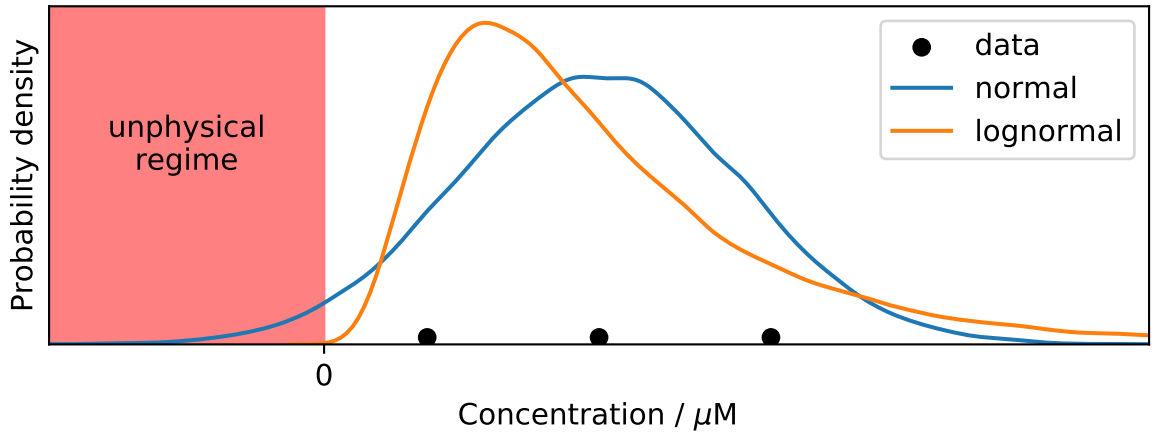


Figure 2.1: Posterior predictive checks for two different likelihood functions describing protein concentrations. Since protein concentrations cannot be negative, values below zero are unphysical.

one does not have to worry if the number of observations is too small. The posterior distribution shows how θ could look like. Bayesian inference will give its most realistic assessment of the distribution of θ. One might decide to collect more observations, e.g. through additional experiments, to narrow it down, though. In contrast, if one wants to fit parameters, one always has to keep in mind that the results might change, perhaps even drastically, if new observations are gathered.

The small size of the available data sets is a common problem in chemistry and biology since data collection is often only possible through laborious and expensive experiments. Therefore, two projects (section 3.2, section 3.3) in this thesis rely heavily on Bayesian statistics. Moreover, section 3.2 demonstrates how multiple data sets can be integrated into a single model using Bayesian inference and how this can be leveraged to gain more insights.

2.1.1 Model Design and Validation

One of the main steps in Bayesian inference is the definition of a probabilistic model that can describe the observed data appropriately. The model links the parameters to the data. Models can be descriptions or approximations of physical processes. They can be defined freely.

For example, if the data are measurements of a molecule's concentration in a sample, these are subject to measurement errors and inaccuracies. A simple model could be to assume that D comes from a normal distribution with mean μ, which would be the real concentration in the sample, and standard deviation σ.

$$D \sim N(\mu, \sigma) \tag{2.4}$$

Following Bayes theorem, a probability distribution for the parameters θ=$\{\mu, \sigma\}$ can be inferred. Posterior predictive checks allow to confirm the validity of the model. Based on θ, new data points can be simulated. These can be compared to the original D. In the example in Figure 2.1, we see that the model generates negative values. These are unphysical since concentrations cannot be negative.

Using a lognormal distribution

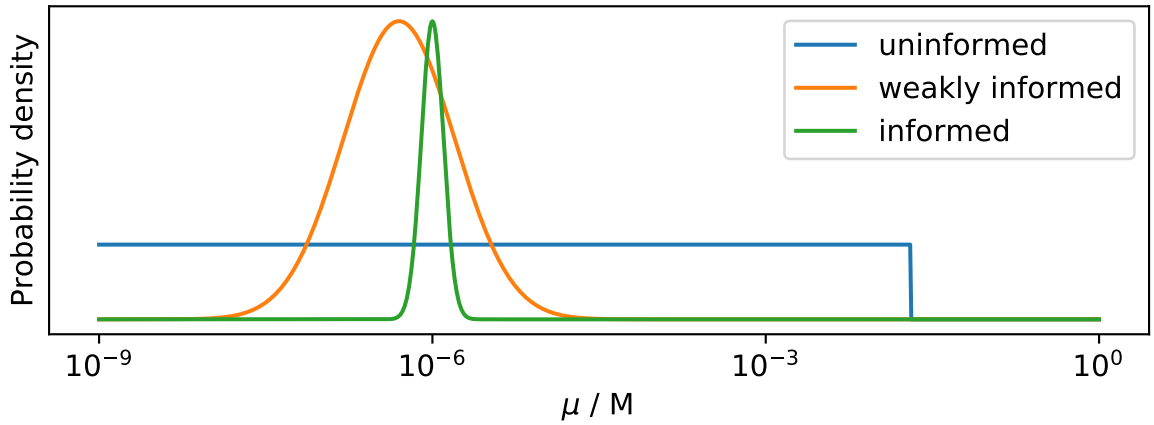


Figure 2.2: Priors for μ, a parameter of the lognormal distribution, which represents the mean of the logarithm of the protein concentration. Different priors contain different amounts of information about the parameter.

$$D \sim \text{Lognormal}(\mu, \sigma) \tag{2.5}$$

instead, negative values are impossible (Figure 2.1). Hence, this likelihood function is more appropriate for the data at hand.

2.1.2 Priors

There are three types of priors. An uninformative prior expresses vague information. In the example considering concentrations, such a prior for μ of the Lognormal distribution could be a simple step function that imposes an upper limit of 20 mM on the concentration (blue line in Figure 2.2). This restriction is motivated by our knowledge that crowded environments, such as the cytoplasm, contain up to 400 g/L macromolecules. [27] Assuming that the molecule we are studying has a molecular weight of 20 kDa (similar to the one of trypsin), the largest concentration to be expected would be 20 mM. Such an uniformed prior would prevent implausible parameters, yet it does not contribute detailed information about the sample at hand.

A weakly informative prior could take into account that similar samples usually contain high nano- or low micromolar concentrations of the protein. It would constrain the parameter to a relatively broad region (orange line). This would make values below 100 nM or above 10 µM unlikely.

Informative priors could be based on detailed knowledge obtained through previous experiments. If the concentration of many similar sample has been found to be 1 µM, and a similar concentration is to be expected in the present sample, one might define a narrow, informative prior (green line). Values that are not close to 1 µM are considered unlikely.

2.1.3 Analytical Inference or Markov Chain Monte Carlo

The adoption of Bayesian statistics has been delayed by the need for computational power to solve Bayes' rule (Equation 2.1) in more complex cases. Ideally, a closed form solution exists. This is sometimes the case if the posterior distribution is in the same probability

distribution family as the prior distribution. Then, the posterior distribution can be determined analytically. Otherwise, the posterior distribution must be approximated. This is done by sampling the distribution using Markov chain Monte Carlo (MCMC). The Metropolis-Hastings algorithm is well known and easy to implement. For complex posterior landscapes, common in Bayesian inference with a multidimensional parameter space, it may struggle to explore the space in a reasonable amount of time. Modern Hamiltonian MCMC algorithms can sample the posterior landscape more efficiently. Bayesian inference packages such as STAN [28] or PyMC3 [29] can provide the analytical gradients of the posterior landscape required for these algorithms.

While these packages are well maintained and offer a plethora of likelihood functions, combining their likelihood functions with a non-conventional, independent model outside of their scope, for example a kinetic model, is not straightforward. In these cases, it may be easier to use a standalone sampler. Unfortunately, analytical gradients cannot be determined for a model that is deemed a black box by the sampler. Numerical determination of the gradients is often computationally expensive. Ensemble samplers, such as the one proposed by Goodman and Weare, [30] do not require gradients but still provide decent performance.

For many models, analytical gradients could be derived *per se*. It may however be simpler and faster to work with numerical implementations of those models, especially if these are already available.

2.1.4 Shortcomings of Null-Hypothesis Significance Testing

During the analysis of a classifier (see section 1.2), one of the central questions is to determine if it is indeed better than random guessing. Bayesian statistics and frequentist statistics follow different interpretations of probability and thus approach this task differently. While we present our Bayesian model in detail in section 3.3, frequentists would resort to null-hypothesis significance testing (NHST). In this scenario, assuming that the classifier is not better than random guessing would be considered the null-hypothesis. NHST allows to determine a p-value. In practice, the null hypothesis is rejected if the p-value is below a threshold, usually 0.05.

Over the last years, opposition to this approach has grown. [31, 32] The extensive use of p-values has been linked to poor reproducibility of scientific studies. [33, 34] Wasserstein summarized the shortcomings of p-values and NHST and how they are often misinterpreted. [32] A p-value is the probability, assuming fulfillment of the null-hypothesis, that the outcome of a study would be equal or more extreme than the observed one. Therefore, p-values can indicate a degree of incompatibility with the null-hypothesis. Yet, among other things, there are the following limitations:

- p-values do not correspond to the probability that the data are the product of randomness alone.
- The conventional 0.05 threshold, separating "statistically significant" from "not statistically significant" findings, is arbitrary. A classifier does not become "better than random" if the p-value decreases below 0.05 nor "random" if it is higher. Reality is more nuanced.
- NHST typically does not measure the size of an effect. If a classifier is better than random guessing, we need to know by how much. For example, if only one in 100 labels are better than random, it is technically better than guessing and the null-hypothesis would eventually be rejected, yet for most practical applications this classifier would still be useless.

- The null-hypothesis is rarely correct. It is unlikely that a classifier's predictions are perfectly random. If the null-hypothesis can be assumed to be false, testing its validity is often not meaningful.

Perhaps the most important step to overcome the problems surrounding the problems with NHST is to accept uncertainty. [35] Data will remain limited in fields where experiments are tedious and expensive. In many cases, one cannot definitely answer the question if a classifier is better than random guessing or not without collecting further data. Often, more nuanced answers to the question "Is this classifier better than random guessing?" are necessary. Rather than a clear "yes" or "no", the answer for many classifiers might be "Perhaps" or "Probably, but we do not know by how much".

Chapter 3

Publications

3.1 Accelerated trypsin autolysis by affinity polymer templates

This section is based on the following publication:

[36] Smolin, D, Tötsch, N, Grad, J.-N, Linders, J, Kaschani, F, Kaiser, M, Kirsch, M, Hoffmann, D, & Schrader, T. (2020) Accelerated trypsin autolysis by affinity polymer templates. *RSC Adv.* **10**, 28711–28719
https://doi.org/10.1039/d0ra05827k

RSC Advances

PAPER

View Article Online
View Journal | View Issue

Cite this: *RSC Adv.*, 2020, **10**, 28711

Accelerated trypsin autolysis by affinity polymer templates†

Daniel Smolin,[a] Niklas Tötsch,[b] Jean-Noël Grad,[b] Jürgen Linders,[a] Farnusch Kaschani,[b] Markus Kaiser,[b] Michael Kirsch,[*c] Daniel Hoffmann[*b] and Thomas Schrader[*a]

Self-cleavage of proteins is an important natural process that is difficult to control externally. Recently a new mechanism for the accelerated autolysis of trypsin was discovered involving polyanionic template polymers; however it relies on unspecific interactions and is inactive at elevated salt loads. We have now developed affinity copolymers that bind to the surface of proteases by specific recognition of selected amino acid residues. These are highly efficient trypsin inhibitors with low nanomolar IC_{50} levels and operate at physiological conditions. In this manuscript we show how these affinity copolymers employ the new mechanism of polymer-assisted self-digest (PAS) and act as a template for multiple protease molecules. Their elevated local concentration leads to accelerated autolysis on the accessible surface area and shields complexed areas. The resulting extremely efficient trypsin inhibition was studied by SDS-PAGE, gel filtration, CD, CZE and ESI-MS. We also present a simple theoretical model that simulates most experimental findings and confirms them as a result of multivalency and efficient reversible templating. For the first time, mass spectrometric kinetic analysis of the released peptide fragments gives deeper insight into the underlying mechanism and reveals that polymer-bound trypsin cleaves much more rapidly with low specificity at predominantly uncomplexed surface areas.

Received 3rd July 2020
Accepted 20th July 2020

DOI: 10.1039/d0ra05827k

rsc.li/rsc-advances

Introduction

Protein autolysis – a natural process difficult to control

Most exocrine proteases undergo autolysis – they digest themselves. Prominent examples are trypsin and chymotrypsin, but also alkaline and neutral proteases exhibit a significant autolysis rate especially at elevated temperatures. Under physiological conditions, autolysis may be crucial for their biological function, but it often hampers their use in biotechnology.[1]

External control of enzymatic self-digest in both directions (inhibition or acceleration) is not easily achieved without side-effects: early work by Hatate and Toyomizu reported accelerated autolysis of trypsin after exposure to polymeric oxidized methyl linolenate, but demonstrated that first ε-amino groups formed Schiff bases with carbonyl groups from the polymer producing an enzymatically inactive covalent polymer/protein adduct.[2] A somewhat related template effect was recently discovered for CAPN3, a protease in skeletal muscles with unusually rapid autolytic activity. When the autolytic N- and C-terminal fragment form a noncovalent complex, they restore proteolytic function (intermolecular complementation iMOC).[3]

Neutral proteases are inactivated at higher temperatures because of autolysis, which involves local unfolding of specific solvent-exposed regions. These could be suppressed by surface-located mutations of N-terminal loops in a neutral zinc metalloprotease.[4] Recently, a stabilization effect for an alkaline serine protease was observed caused by copper ions (5 mM) as the result of a decrease in both autolysis and thermoinactivation rates.[5] Autolysis has been identified as the primary mode of subtilisin activity loss in a heavy-duty liquid detergent (HDLD) formulation; it could in part be repaired by thermodynamic stabilization and/or kinetic inhibition.[6] In general, enzymes can be protected from protease digestion by modification with hydrophilic well-hydrated polymers, due to their shielding effects.[7] A very interesting contribution along these lines came from Sasai *et al.*, who synthesized highly stabilized polymer–trypsin conjugates with remarkable autolysis resistance. Thus, a vinylmethylether–maleic acid copolymer (VEMAC) was covalently bound to trypsin's lysine amino groups *via* multi-point attachment. The modified trypsin construct showed strong resistance against autolysis and unaltered conformation (CD spectra), explained by mutual electrostatic repulsion of the negatively charged surfaces of the VEMAC-coated trypsin.[8]

[a]*Faculty of Chemistry, University of Duisburg-Essen, 45117 Essen, Germany. E-mail: thomas.schrader@uni-due.de*

[b]*Faculty of Biology, University of Duisburg-Essen, 45117 Essen, Germany*

[c]*University Clinics, Essen, Germany*

† Electronic supplementary information (ESI) available: Materials and methods, enzyme assays, fluorescence and ITC titrations, CD spectroscopy, DLS, gel filtration, CZE, SDS-PAGE, PFG-NMR, ESI-MS fragment analysis, computation. See DOI: 10.1039/d0ra05827k

On the other hand, it may be highly desirable to accelerate autolysis of problematic or even pathogenic proteins. In principle, this may be achieved by increasing their local concentration by assembly on a large template molecule, *e.g.* an oppositely charged polymer. Such an example has recently been presented for the first time by Lv *et al.*, who were able to accelerate trypsin's self-digest in the presence of polyelectrolytes, most prominently dextrane sulfate.[9] In their seminal work the authors demonstrated how direct interaction between trypsin and the negatively charged polymer would greatly diminish the enzymatic activity of the protease, through a rapid self-cleavage into smaller fragments. They called this effect "autolysis-acceleration protease inhibition" (AA-PI) and investigated it with MALDI-TOF, SDS-PAGE, DLS and TEM. From their experiments, the authors deduced that doubly charged monomer units are required, and they also found that, using such polymers, insulin could be protected against trypsin degradation, but only at low ionic strength. Unfortunately, these polymers completely lose their inhibition ability under elevated salt loads, *i.e.* under physiological conditions, most likely because they primarily rely on multiple electrostatic attraction. At the end of their account, Lv *et al.*[9] conclude that more specific AA-PIs are necessary, and that these may be able to inhibit deleterious pathogenic proteases.

Designed affinity copolymers as specific AA-PIs

We have developed two such specific AA-PIs by supramolecular chemistry. They inhibit their target protease in a highly efficient manner, and do not lose their activity under elevated salt loads. Among the most prominent digestion enzymes they are specific against trypsin and the closely related kallikrein, which they inhibit in a drastically substoichiometric manner. They are part of our project to develop protein-selective copolymers from amino-acid-selective monomer units. In this case, we developed two binding monomers tailored for lysine/arginine and serine, the two most abundant amino acids on the trypsin surface. These recognize their targets differently from simple electrolytes (Fig. 1). A lysine or arginine residue, *e.g.*, is noncovalently bound by the bisphosphonate monomer with a combination of noncovalent interactions, including hydrogen bond reinforced Coulomb attraction, but also π-cation and dispersive interactions as well as the hydrophobic effect (displacement of high energy water). Serine, on the other hand is covalently bound by ester formation with aminomethylphenylboronic acid. Linear copolymerization of one or both co-monomers in the absence of the protein target leads to water-soluble copolymers that are able to complex the protein surface of trypsin by multiple (non) covalent interactions. Since trypsin's active site is not directly flanked with lysines, arginines or serines, its proteolytic activity is not compromised by this complexation event. Up to 10 copies of the serine protease can be accommodated on the linear copolymer (Fig. S11†), leading to a high local concentration that accelerates mutual self-digest. In this paper, we discuss the features of this polymer-assisted self-digest, present a computational model and provide deeper insight into the putative mechanism by proteomics MS analysis.

The above-mentioned two trypsin-specific polymers represent a larger family of affinity copolymers that are able to engage in multipoint binding[10] to hot spots on proteins: they rely on the construction of a series of specific binding monomers for each class of amino acids. Those that are complementary to critical residues on the hot spot, are mixed in defined stoichiometric ratios and subjected to radical copolymerization. The resulting linear copolymers display high affinity and surprising protein specificity by cooperative binding.[11] Very recently we developed by this method highly specific polymeric inhibitors for disease-relevant proteases.[12] Their specificity comes from the multivalent combination of well-defined binding motifs for critical amino acids on the respective protein surface.[13]

Results and discussion

Synthesis and properties of P1 and P2

Two powerful polymeric trypsin binders were synthesized from their respective methacrylamide comonomers equipped with a bisphosphonate anion[10] and an aminomethylphenylboronic acid moiety.[12] Polymerization was started by AIBN and conducted for 24 h in water/DMF. After lyophilization, the crude products were redissolved in ultrapure water and subjected to ultrafiltration to remove monomers and oligomers. Final lyophilization furnished colorless powders with excellent water solubility (yields 55–60%). NMR integration confirmed that the original stoichiometric monomer ratio was retained in **P2**. GPC produced average molecular weights for both polymers of ~170 kD with relatively low polydispersities of 1.4–1.5, most likely due to the ultrafiltration cut-off. These molecular weights correspond to a degree of polymerisation DP of 600 for **P1** and 730 for **P2**. The copolymerization parameters were determined earlier for the bisphosphonate monomer and a related polar comonomer with a secondary alcohol. In water, r_1 and r_2 remained between 0.3 and 2.0 (Fineman–Ross), while in DMF they became 1.0, indicating ideal statistical copolymerization.

Tailored affinity polymers P1/P2 accelerate trypsin self-digestion

When we studied the two new trypsin inhibitors with multiple arginine/lysine and serine binding sites (**P1**: bisphosphonate homopolymer, **P2**: bisphosphonate/boronate copolymer), we discovered some unusual features. Contrary to most other polymers, including the naturally occurring polypeptides like aprotinin (1 : 1), IC_{50} values dropped to drastically substoichiometric ratios, indicating a 1 : 10 (**P2**) or even 1 : 100 (**P1**) polymer:protein stoichiometry (see ESI†). In addition, prolonged incubation of enzyme and polymer further lowered IC_{50} values, consistent with a so-called slow-onset mechanism (see ESI†).[14] These initial observations cannot be explained with conventional substrate–polymer competition, but agree very well with the assumed template effect of the affinity polymers for enhanced enzymatic self-cleavage.

First evidence for the postulated polymer accelerated self-digest (PAS) came from gel electrophoresis of the trypsin cleavage products in the absence and presence of affinity polymers. An

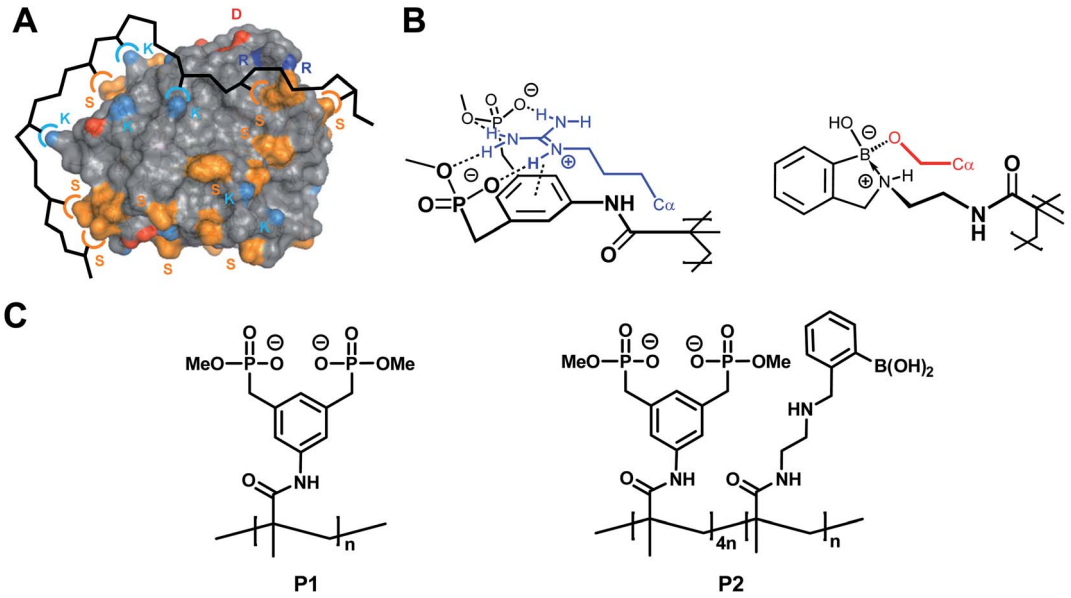


Fig. 1 Concept of multivalent protein surface recognition by linear affinity copolymers from amino acid-selective binding monomers. (A) Schematic of the affinity copolymer concept – each color symbol represents an amino acid/binding monomer pair. (B) Molecular recognition pattern for the specific targeting of arginine (left) and serine residues (right) on the protein surface by designed binding sites in affinity copolymers. Note the noncovalent combination of ionic hydrogen bonds and π-cation attraction between bisphosphonate dianion and alkylguanidinium side chain (blue), and reversible covalent formation of a cyclic boronate ester between the aminomethylphenylboronic acid moiety and the primary alcohol (red). (C) Lewis structures of homopolymer P1 and copolymer P2.

incubation period of trypsin with **P1** or **P2** in Tris buffer (37 °C) in the absence of any enzyme substrates was terminated at selected time points by the addition of SDS PAGE loading buffer containing dithiothreitol (DTT), and the integrity of the protein was analyzed by SDS-PAGE (Fig. 2A).[15] While native trypsin needed 4 h for complete self-digest (all bands), this stage was already reached after 10–15 min in the presence of 0.1 equiv. of **P1** or **P2**. The largest bands appear to be cleaved faster in the polymer-catalyzed case, leaving a slightly altered molecular weight distribution compared to trypsin alone. The smallest fragments accumulated at the bottom of the gel indicating complete proteolytic destruction into small peptides. When time-dependent CD spectra were recorded at identical protein/polymer ratios (37 °C), the observed effects from gel electrophoresis were strongly supported (Fig. 2B): trypsin self-digest in the absence of polymer needed more than 30 min to dissolve the typical CD band at 212 nm; with **P1** or **P2**, however, this band disappeared already after 5 min (for full CD kinetics see ESI†).

Gel filtration is a powerful method to monitor the molecular weight distribution of proteins;[16] we wondered if it could be utilized to demonstrate trypsin self-cleavage into smaller fragments in a time-dependent manner. Our first attempts failed because virtually no changes occurred in the initial chromatogram under the established conditions for catalyzed self-digest (60 μM trypsin, 6.7 μM **P1**, 100 mM borate buffer, 37 °C). However, in the presence of both DTT (10 mM) which reduces disulfide bonds[17] – and a denaturing agent (1 M urea), the desired changes became visible (Fig. 2C): trypsin bands at 18 mL (22 kDa) and 20–23 mL elution volume (5–17 kDa) decreased rapidly only in the presence of **P1** or **P2** (already after 5 min), and furnished new increasing autolysis bands at 23–28 mL corresponding to molecular weights between 0.5 and 1 kDa. As a control, the affinity polymer band was constant at 11–15 mL (30–100 kDa), and DTT likewise appeared constant at 30 mL (0.3 kDa). Molecular weights were estimated from retention volumes of reference proteins. Intriguingly, in the polymer-catalyzed system, the largest band (β-trypsin)[18] was cleaved much more rapidly than the smaller bands. In the growing product bands, molecular weights were gradually shifted towards smaller fragments, illustrating the constant degradation even of medium sized protein fragments on the polymer template.

Another method for chromatographic separation of proteins from polymers is capillary zone electrophoresis CZE.[19] Under the above-described standard conditions, native trypsin reproducibly elutes as a typical mixture of large bands with a characteristic ratio (see ESI†). In the presence of very small amounts of **P1** (0.5 mol%) and **P2** (0.8 mol%), however, trypsin was fully degraded after less than 20 min, whereas self-digest in their absence remained slow.[20] The stoichiometric polymer/protein ratio for these experiments leading to complete trypsin degradation reached 1 : 185. Parallel enzyme assays confirmed the loss of 95% of the original trypsin activity. These values place **P1** and **P2** among the most potent trypsin inhibitors known to date.[21]

Detailed analysis of the inhibition mechanism and the underlying protein polymer recognition event

For a better understanding we investigated the complex formation between trypsin and the polymers more closely.

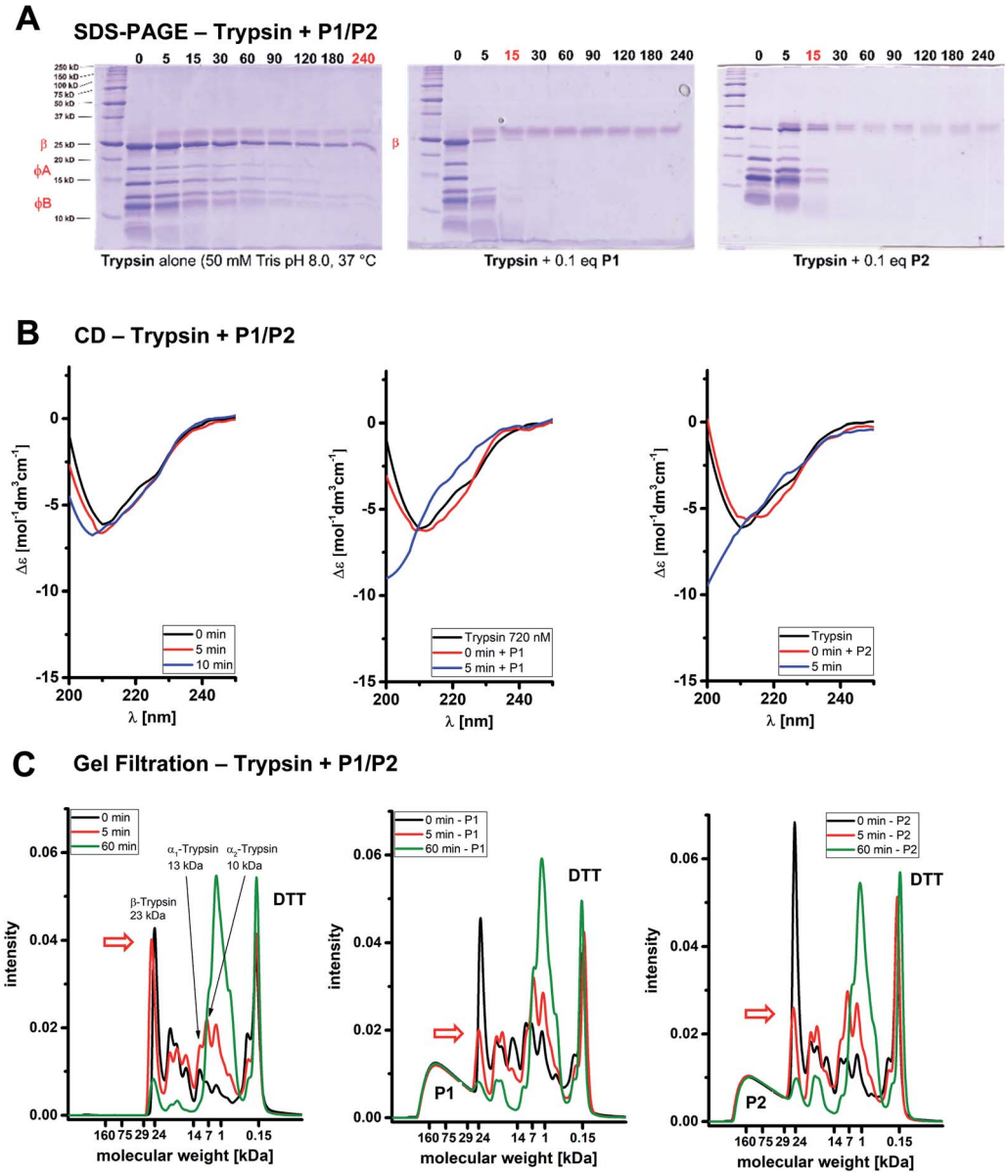


Fig. 2 Affinity polymers **P1/P2** accelerate trypsin self-digestion. (A) SDS-PAGE. Autolysis of 60 μM trypsin over time (75 mM TRIS buffer, pH 8.0, 37 °C). Left: Trypsin alone; right: trypsin with 6.7 μM **P1/P2**. (B) CD spectroscopy. Time-dependent CD spectra monitoring the kinetics of polymer-accelerated trypsin autolysis (60 μM trypsin, 75 mM borate buffer). Left: Trypsin alone; right: trypsin with 6.7 μM **P1/P2**. (C) Gel filtration. Gel filtration chromatograms depicting the molecular weight distribution of trypsin alone and in the presence of **P1** during the self-digest process (100 mM borate buffer, denaturated with 6 M urea and 30 mM DTT). Left: Trypsin alone, right: trypsin with **P1/P2**. **P1/P2** concentrations refer to the full polymer molecular weight.

Substoichiometric and slow onset inhibition prevailed in various buffers, as long as the pH was kept slightly acidic or slightly basic (pH 6–9). Preincubation at acidic pH < 3 completely abolished all effects. Under these conditions, self-digest does not occur because trypsin activity is greatly reduced if the pH is below 7 or above 9. Moreover, all bisphosphonate units become protonated and lose their affinity for basic amino acid residues.[22] In neutral buffer, however, even at high ionic strength such as 150 mM PBS the inhibition efficiency of **P1** and **P2** is not lowered, which points to a powerful combination of Coulomb attraction with other (non)covalent interactions, a built-in feature of our amino acid-selective binding monomers. This contrasts with the performance of non-specifically binding polyelectrolytes such as dextrane sulfate, which largely rely on multiple electrostatic attraction, and completely lose their inhibition ability under elevated salt loads.[9]

Attempted recovery experiments with cationic polymers were frustrated according to both CD spectra as well as enzyme activity assays: irreversible inhibition occurred in various buffers; in other words, the enzyme did not return to its active native state even when the polymer is completely peeled off its

surface (see ESI†). Such a reversible denaturation was observed when lysozyme was bound and inhibited by a specific affinity polymer, followed by addition of polyarginine, which turned on enzymatic activity again.[23] These experiments rule out a conventional inhibition mechanism that involves reversible coverage of the active site.

In order to exclude unspecific effects from aggregate formation, dynamic light scattering (DLS)[24] was performed on trypsinogen (no self-digest) and the isolated polyanionic polymers: monodisperse hydrodynamic radii of 2.4 nm for trypsinogen (lit.: 2.2 nm)[25] and 10 nm for both polymers **P1**/**P2** are in good agreement with single protein and polymer molecules in a non-aggregated state. Further confirmation comes from GPC determination of polymer molecular weights at $M_n \approx 100$ kDa. Intriguingly, the hydrodynamic radius of **P2** decreased from 10 to ≈7 nm when it was complexed with trypsinogen, indicating a more tightly folded state of the polymer (see ESI†).

Very similar results came from PFG-NMR (Pulsed Field Gradient), which produced the hydrodynamic radii of 1 nm for trypsin and 10 nm for the polyanionic polymers, strongly suggesting that polymer and protein did not form aggregates.[26] PFG-NMR measurements also were consistent with a drop of polymer radius to about 70% on complex formation with trypsin (see ESI†).[27] This effect was independently predicted by a computer model presented below. Fluorescence titrations only led to moderate quenching; for **P1** a binding isotherm could be evaluated quantitatively by nonlinear regression and furnished a moderate K_d value of 540 nM at a 9 : 1 stoichiometry (see ESI†).[10] This is very important, because it reveals for the first time that 9 trypsin molecules can be accommodated on one polymer strain simultaneously.[28] Since both binding partners carry multiple and opposite charges, electrostatic attraction will most likely be maximized by a substantial folding of the polyanionic polymer around its 9 polycationic trypsin guests leading to a collapsed complex state, as evidenced by DLS and PFG-NMR. Moderate polymer protein affinity enables dissociation of large cleavage products from the complex and secures catalytic turnover, but limits catalytic efficiency. These features are also discussed in the simulation below.

The new polymer-accelerated self-digest can be simulated by a simple theoretical model

In line with the above-discussed experimental observations we propose the following mechanistic model which correlates well with that from Lv *et al.*:[9] Polymer Accelerated Self-digest (PAS) is a process in which a polymer in aqueous solution accelerates the self-digest of a protease. Basic ingredients of PAS are a protease that is cleavable by its own kind, and a flexible polymer with affinity to this protease, though not blocking its active site. The polymer has to be long enough to bind several protease molecules. Fig. 3A illustrates the PAS mechanism schematically for the protease trypsin.[29] In the solution of polymer and protease (a) a polymer molecule binds several protease molecules (b), often in a multivalent fashion, leading to formation of loops and to polymer collapse. This interaction between polymer and protein concentrates protease molecules in a small volume, causing efficient cannibalistic autolysis. (c) This cannibalistic self-digest is probably incomplete because it requires a small amount of functional protease molecules until the end. (d) The more flexible digest peptides compete with intact, well-structured protease molecules that are still in solution. The result of this competition is likely a release of the digested peptides, which completes the catalytic cycle and allows the polymer to bind further protease molecules. (e) Protease binding, autolysis, and digest release proceed continuously as long as a sufficient number of protease molecules are available in the solution.

This mechanism is a beautiful example of the powerful influence of multivalency on biological activity. Only the synergistic binding of many enzyme molecules on the same polymer template leads to the observed high local concentrations and enables the polymer to drastically accelerate trypsin autolysis. A visualization of a stabilized trypsin dimer is given in Fig. 3B (trypsin from PDB 2ptn). These multiple protein–polymer contacts must also be reversible and transient to maintain a catalytic cycle where cleaved peptide fragments dissociate from the polymer template.[30] If the PAS mechanism sketched above is correct, we should be able to recapitulate major elements of this mechanism with computational models that share essential features with the real system.

We used epitopsy[31] to computationally screen the surface of trypsin with BP monomers, the main building block of our polymers. The analysis shows several patches on trypsin that have elevated affinity for BP, though the active site of trypsin is not covered (Fig. 3C). This is exactly the pattern expected for a protein that binds multivalently to a BP-based polymer, while the polymer spares the active site region and thus does not inhibit proteolytic activity.

A pattern as the one in Fig. 3C is necessary for PAS but it only accounts for interactions of monomers with a single trypsin molecule. Other observed features of PAS cannot be understood at this level, *e.g.* the collapse of the polymer in the presence of trypsin molecules, and of course the highly efficient self-digest of trypsin in the presence of the polymer. These aspects have to be addressed with models comprising at least a complete polymer molecule and several trypsin molecules. Since the system of polymer and trypsin molecules in aqueous solution is complex and has a huge configurational space, we opted for a minimalist lattice model that allows for a more thorough sampling of configurational space.[32] This lattice model consists of a flexible polymer model surrounded by a number of trypsin molecules, which can bind multivalently to the polymer, with each polymer–trypsin contact contributing a fixed free energy of binding ("contact energy" E_{cont}).

We report here some of the key results from the lattice model simulations (Fig. 3D). These results are in good agreement with experiment and support the PAS model. First, the simulations predict that the interaction with trypsin molecules leads to polymer collapse: the mean monomer–monomer distance R_m, measured as fraction of the mean monomer–monomer distance of the free polymer, drops significantly in the presence of trypsin (Fig. 3D(A)). The drop of R_m becomes more drastic with increasing degree of polymerization (DP), from $R_m = 0.84$ at DP

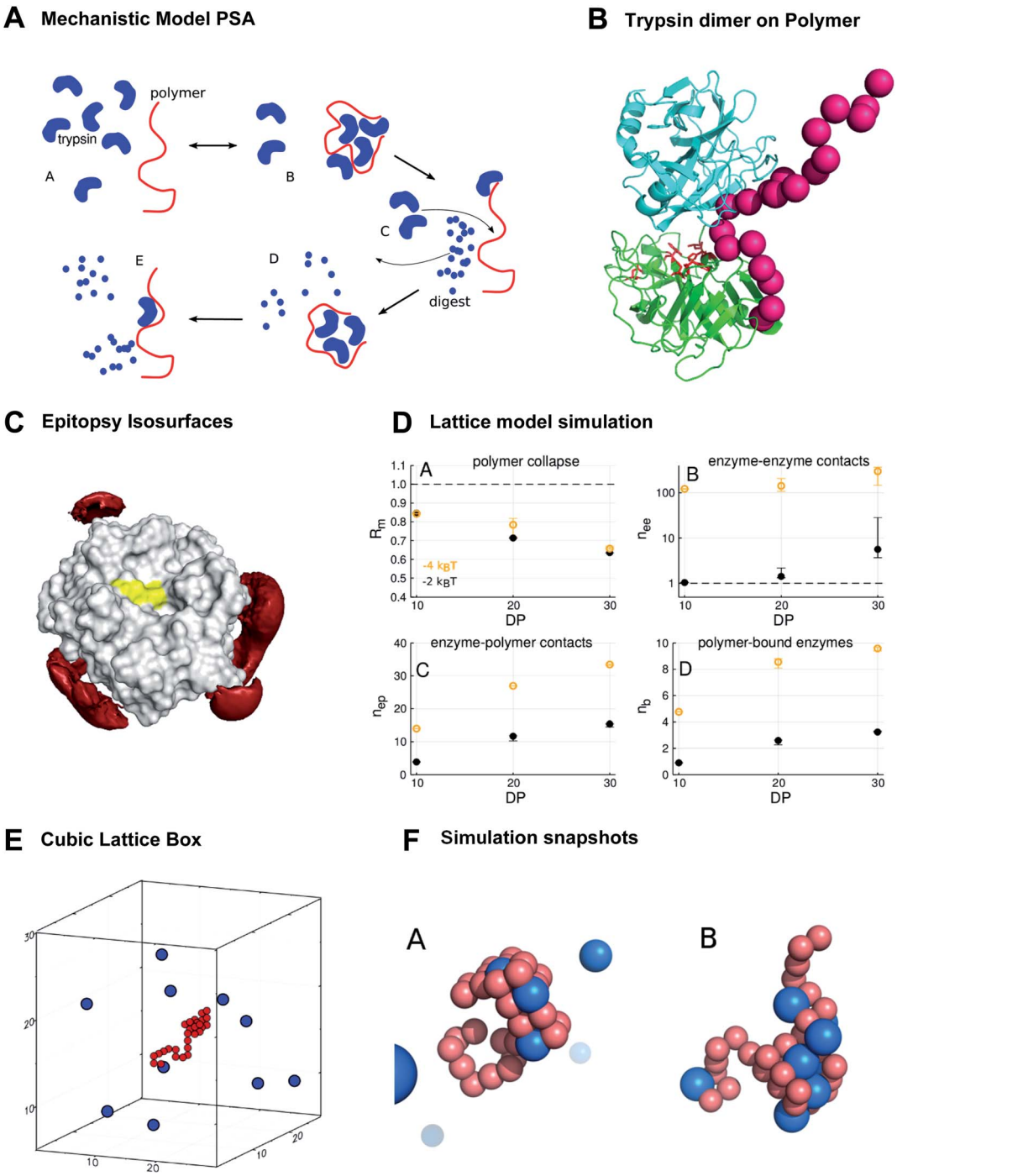


Fig. 3 The new polymer-accelerated self-digest can be simulated by a simple computational model. (A) Proposed mechanistic model starting with intact protease and polymer (A), followed by polymer collapse (B), cannibalistic autolysis (C) and released peptide fragments (D and E). (B) Polymer (red) stabilizes a trypsin dimer (blue, green), which favors cannibalistic autolysis (active center in red sticks). (C) Red isosurfaces of increased affinity level ($-1k_BT$) between trypsin and the bisphosphonate monomer, which binds to K/R without shielding the active site (yellow). (D) Lattice model simulation results. DP = degree of polymerization. Polymer contact $-2kT/-4kT$. (A) Mean monomer–monomer distance R_m; (B) average number n_{ee} of enzyme–enzyme contacts; (C) average number n_{ep} of enzyme–polymer contacts; (D) average number n_b of enzymes bound to polymer (total enzymes = 10). (E) Cubic lattice box with 10 enzymes and polymer of DP.[30] (F) Simulation snapshots. (A) 30-mer polymer (red) interacts with trypsin molecules (blue) at contact energies $E_{cont} = -2k_BT$ (A) and $E_{cont} = -4k_BT$ (B) at the same trypsin concentration. Only in (B) enzymes contact each other, illustrating the importance of strong single interactions.

= 10 to about $R_m = 0.65$ at DP = 30, which is close to the experimentally observed collapse to 70% of the polymer radius on interaction with trypsin (see ESI†).

Interestingly, it does not need many polymer-bound proteases for a polymer collapse. Even if we have on average only a single protease molecule bound to the polymer molecule with

a weak contact energy, as in the case of DP = 10 and $E_{cont} = -2k_BT$ (point in lower left corner of Fig. 3D(D)), the value of R_m drops to 0.84 (point in upper left corner of Fig. 3D(A)). This is because protease molecules in the model are typically bound to the polymer multivalently with on average 3 to 5 contacts per protease molecule (compare Fig. 3D(C and D)). Fig. 3F illustrates this with two simulation snapshots. Thus, multivalent protease–polymer binding stabilizes loops and coiling of the polymer and therefore leads to the observed collapse.

The crucial feature of PAS – that is supported by the lattice simulations – is the enrichment of protease–protease contacts in the presence of the polymer (Fig. 3D(B)). The polymer collects proteases from the solution (Fig. 3D(D)) and fixes them by multivalent binding (Fig. 3D(C)) in a small fraction of the available volume. Thus, in comparison to the polymer-free solution, the number n_{ee} (with index ee for "enzyme–enzyme") of protease–protease contacts can increase by orders of magnitude, depending slightly on DP, but strongly on protease–polymer affinity (Fig. 3D(B)).

The increase of n_{ee} with DP is essentially due to the higher abundance of binding sites in larger polymers. The increase of n_{ee} with protease–polymer affinity (in Fig. 3D(B) shift from black points, $-2k_BT$, to orange points, $-4k_BT$) can be attributed mainly to facilitated protein capture by higher affinity binding sites on the polymer.

However, there is also a more subtle effect that contributes to the upward shift of n_{ee} for stronger protease–polymer affinity: for weaker binding energy E_{cont} per protease–polymer contact, protease–polymer complexes with single contacts are thermodynamically unstable. Therefore, stable binding conformations that we observe under these conditions will typically involve per protease several contacts to the polymer. Thus, the polymer will occupy a substantial fraction of the surface of each bound protease, and these regions will then not be available for contacts to other proteases – the polymer partly shields proteases against each other (Fig. 3F(A)). Conversely, for stronger protease–polymer contact energies E_{cont}, even single polymer–protease contacts may be thermodynamically stable, enabling a bound protease to interact with other proteases (Fig. 3F(B)). In our simulations we found that at a contact energy of $-2k_BT$, a protease molecule was on average engaged in 4.3 to 4.7 contacts with the polymer, while at $-4k_BT$ the numbers were lower at 2.9 to 3.5, leaving more opportunities for protease–protease contacts.

Protease–polymer aggregation, a key aspect of the above-postulated mechanism, can be observed with the naked eye at elevated concentrations (see ESI†): addition of trypsin to 0.1 equiv. **P1** results in instant precipitation of the polymer/protein complex. During the following 15 min, however, the solution becomes gradually less cloudy due to trypsin degradation, finally leading again to a clear solution. Another addition of trypsin repeats the whole cycle until a clear polymer solution is again obtained. For an efficient turnover it is very important, that the trypsin fragments leave the polymer template in order to avoid product inhibition or poisoning of the catalyst.[33]

A somewhat related principle was recently used to promote B and T cell activation by a polymeric bifunctional multivalent

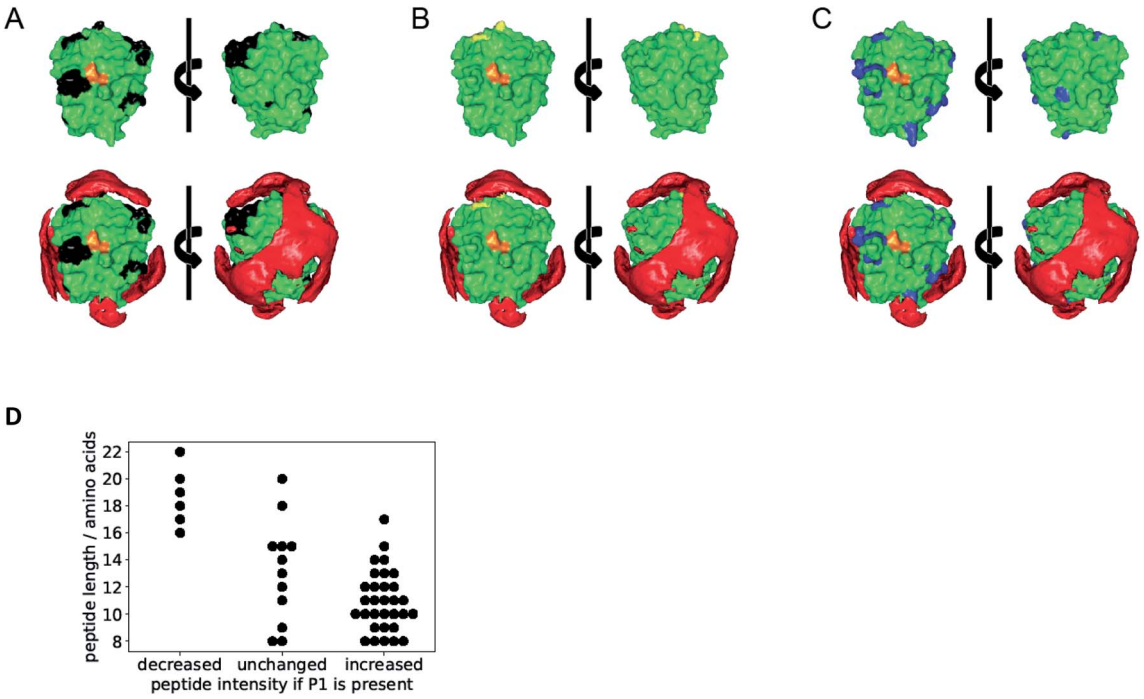


Fig. 4 Overlay of epitopsy calculation with cleavage sites identified by MS. The active site of the enzyme is depicted in orange and represents the front of the enzyme. (A) Black cleavage sites found in peptides produced with increasing concentration upon treatment with **P1**. (B) Yellow cleavage sites from peptides with decreased concentration upon treatment with **P1**. (C) Blue peptides with no significant abundance change upon **P1** treatment. (D) Accelerated autolysis leads to shorter peptide fragments. Each point corresponds to one specific digest peptide. Presence of **P1** decreases abundance of longer peptides and increases abundance of shorter peptides.

antigen.[34] Here, the multivalent presentation of the B cell epitope (DNP) on a ROMP polymer led to efficient oligomerization of the BCR, which in turn induced both signaling and uptake. Similar to our case with PAS, a synthetic polymer acts as a multivalent template to bring about a biologically significant oligomerization event.

Mass spectrometry: trypsin bound to polymers P1/P2 cleaves with low specificity predominantly at uncomplexed surface areas

For deeper insight into the templated cleavage process triggered by the addition of the polymer **P1** we performed a mass spectrometric analysis of the released peptide fragments. In a bottom-up approach we initiated trypsin self-digest under optimized conditions (borate buffer pH 8.0) in the presence and absence of **P1** and at varying reaction times. The reactions were terminated (5% formic acid) and subsequently, an excess of acetone was added which is known to precipitate undigested proteins and partially digested polypeptides while smaller oligopeptides generated by autolysis at the different time points remain in solution (Fig. S19†).[35] After removing insoluble matter by centrifugation acetone was evaporated and the remaining peptides were analyzed by LC-MS/MS. After stringent filtering 49 specific trypsin degradation peptides were identified, whose intensities were normalized (Fig. S20 and File S20†). Their majority (31 peptides) showed a marked increase in abundance upon treatment with **P1**, a small fraction (6 peptides) showed partially decreased abundance, and 12 peptides showed no marked effect.

Interestingly, the surface-exposed cleavage sites from peptides with increased or unchanged abundance upon **P1** treatment are predominantly in regions with low affinity towards BP according to the epitopsy calculation and therefore expected to remain accessible if **P1** is present (Fig. 4A–C). The agreement between MS and epitopsy is encouraging although there are major approximations (see sections on MS and epitopsy in the ESI and Fig. S21†). **P1** apparently increases local concentration without blocking the active site or covering too many peptide bonds susceptible to proteolysis which is a requirement for our suggested mechanism.

Digested peptides decreasing in abundance if **P1** is present were substantially longer than other peptides (Fig. 4D). Thus our results suggest a promotion of trypsin autolysis by **P1**, either by a more thorough digest leading to smaller products, or a faster autolysis, or a combination of both.

Surprisingly many peptides were not fully tryptic (ESI File S20†); most identified peptides were semi-tryptic. We can conclude that in the presence of **P1** cleavage specificity was low. The reason for unspecific trypsin autolysis in the presence of **P1**/**P2** is unclear at present.

Conclusion and outlook

In summary, we have demonstrated how linear affinity copolymers that specifically recognize the protein surface lead to accelerated autolysis. Presumably, they act as a template for multiple protease molecules. Local protease concentration on the accessible surface area is substantially higher than in the bulk solution and complexed areas are shielded from digestion. The extremely efficient trypsin inhibition has been monitored by SDS-PAGE, gel filtration, CD, CZE and ESI-MS. Such a controlled and specific enzyme self-destruction may be helpful to counteract trypsin upregulation in ischemia and reperfusion injuries of the small intestines as well as in acute pancreatitis. It may also be directed against other problematic proteases involved in diseases (*e.g.*, retroviral protease or β-secretase). We have recently discovered that affinity copolymers tailored for kallikrein also exhibit drastically substoichiometric protein inhibition and will likewise examine their mechanism of action supported with calculations. Moreover, controlling protease activity is highly desirable to protect therapeutic peptides. Another potential application involves the accelerated digest of other selected proteins in the presence of a small amount of trypsin and an appropriate affinity polymer. Employing living radical copolymerization, we envisage the construction of block copolymers for simultaneous trypsin and protein substrate docking and accelerated substrate digest. Experiments in this direction are underway in our laboratories.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors gratefully acknowledge financial support from the Deutsche Forschungsgemeinschaft DFG by funding the Collaborative Research Centre CRC 1093 Supramolecular Chemistry on Proteins.

References

1 D. C. Whitcomb, M. Gorry, R. Preston, *et al.*, *Nat. Genet.*, 1996, **14**, 141–145.
2 H. Hatate and M. Toyomizu, *Bull. Jpn. Soc. Sci. Fish.*, 1985, **51**, 627–633.
3 Y. Ono, M. Shindo, N. Doi, F. Kitamura, C. C. Gregorio and H. Sorimachi, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, E5527–E5536.
4 S. A. Shirdel, K. Khajeh, S. M. Asghari and H. R. Karbalaei-Heidari, *Eng. Life Sci.*, 2014, **14**, 229–234.
5 N. C. Ozturk, D. Kazan, A. A. Denizci and A. Erarslan, *Eng. Life Sci.*, 2012, **12**, 662–671.
6 M. R. Stoner, D. A. Dale, P. J. Gualfetti, T. Becker, M. C. Manning, J. F. Carpenter and T. W. Randolph, *Enzyme Microb. Technol.*, 2004, **34**, 114–125.
7 F. M. Veronese and G. Pasut, *Drug Discovery Today*, 2005, **10**, 1451–1458.
8 Y. Sasai, H. Kanno, N. Doi, Y. Yamauchi, M. Kuzuya and S.-I. Kondo, *Catalysts*, 2017, **7**, 4–13.
9 Y. Lv, J. Zhang, H. Wu, S. Zhao, Y. Song, S. Wang, B. Wang, G. Lv and X. Ma, *Chem. Commun.*, 2015, **51**, 5959–5962.

10 C. Renner, J. Piehler and T. Schrader, *J. Am. Chem. Soc.*, 2006, **128**, 620–628.
11 S. Koch, C. Renner, X. Xie and T. Schrader, *Angew. Chem.*, 2006, **38**, 6500–6503.
12 P. Gilles, K. Wenck, I. Stratmann, M. Kirsch, D. A. Smolin, T. Schaller, H. de Groot, A. Kraft and T. Schrader, *Biomacromolecules*, 2017, **18**, 1772–1784.
13 T. Schrader, *Chem.–Eur. J.*, 1997, **3**, 1537–1541.
14 (*a*) A. Baici, Slow-Onset Enzyme Inhibition, *Kinetics of Enzyme-Modifier Interactions*, Springer, Wien, 2015, pp. 367–444; (*b*) H.-J. Li, C.-T. Lai, P. Pan, W. Yu, N. Liu, G. R. Bommineni, M. Garcia-Diaz, C. Simmerling and P. J. Tonge, *ACS Chem. Biol.*, 2014, **9**, 986–993.
15 J. S. Hanspal, G. R. Bushell and P. Ghosh, *Anal. Biochem.*, 1983, **132**, 288–293.
16 P. Hong, S. Koza and E. S. P. Bouvier, *J. Liq. Chromatogr. Relat. Technol.*, 2012, **35**, 2923–2950.
17 U. T. Ruegg and J. Rudinger, *Methods Enzymol.*, 1977, **47**, 111–116.
18 M. Vestling, C. M. Murphy and C. Fenselau, *Anal. Chem.*, 1990, **62**, 2391–2394.
19 Capillary Electrophoresis of Proteins and Peptides, Methods and Protocols, *Methods in Mol. Biology*, ed. N. T. Tran and M. Taverna, Humana Press, Springer, New York, 2016.
20 This somewhat contrasts the 20% rest activity found in enzyme assays; however, CZE does not measure activity but molecule charge and size.
21 It has to be acknowledged that the polymers do not act as classical inhibitors, but rather as catalytic accelerators of tryptic self-digest by means of a template effect.
22 The pK_a value of the bisphosphonic acid was determined at 2.5.
23 K. Wenck, S. Koch, C. Renner, W. Sun and T. Schrader, *J. Am. Chem. Soc.*, 2007, **129**, 16015–16019.
24 (*a*) J. Mogridge, *Methods Mol. Biol.*, 2004, **261**, 113–118; (*b*) K. Gast and C. Fiedler, *Methods Mol. Biol.*, 2012, **896**, 137–161.
25 G. Caracciolo, *et al.*, *Eur. Biophys. J.*, 2001, **30**, 163–170.
26 J. Linders, C. Mayer, T. Sekine and H. Hoffmann, *J. Phys. Chem. B*, 2012, **116**, 11459–11465.
27 M. Tirrell, *ACS Cent. Sci.*, 2018, **4**, 532–533.
28 C. Y. Huang, *Methods Enzymol.*, 1982, **87**, 509–525.
29 J. Walter, W. Steigemann, T. P. Singh, H. Bartunik, W. Bode and R. Huber, *Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem.*, 1982, **38**, 1462–1472.
30 C. Fasting, C. A. Schalley, M. Weber, O. Seitz, S. Hecht, B. Koksch, J. Dernedde, C. Graf, E.-W. Knapp and R. Haag, *Angew. Chem., Int. Ed.*, 2012, **51**, 10472–10498.
31 J.-N. Grad, A. Gigante, C. Wilms, J. N. Dybowski, L. Ohl, C. Ottmann, C. Schmuck and D. Hoffmann, *J. Chem. Inf. Model.*, 2018, **58**, 315–327.
32 (*a*) D. P. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics*, Cambridge University Press, Cambridge, 2009; (*b*) M. N. Rosenbluth and A. W. Rosenbluth, *J. Chem. Phys.*, 1955, **23**, 356–359.
33 This reversible turbidity increase exactly parallels the behaviour of dextrane sulfate in ref. 9.
34 N. R. Bennett, D. B. Zwick, A. H. Courtney and L. L. Kiessling, *ACS Chem. Biol.*, 2015, **10**, 1817–1824.
35 K. Bravo-Rodriguez, B. Hagemeier, L. Drescher, M. Lorenz, J. Rey, M. Meltzer, F. Kaschani, M. Kaiser and M. Ehrmann, *Rapid Commun. Mass Spectrom.*, 2018, **32**, 1659–1667.

Supplementary Information

Accelerated Trypsin Autolysis by Affinity Polymer Templates

Daniel Smolin, Niklas Tötsch, Jean-Noël Grad, Jürgen Linders, Farnusch Kaschani, Markus Kaiser, Michael Kirsch*, Daniel Hoffmann,* Thomas Schrader*

Faculty of Chemistry, University of Duisburg-Essen, 45117 Essen, Germany

Materials and Methods

General Information
All reagents were purchased from Sigma-Aldrich and were used as received. The used water was ultrapure water purified with a ELGA PURELAB Classic UV system.

^{1}H and ^{31}P NMR spectroscopy were performed on a Bruker DMX300 or Bruker DRX500, as specified. Chemical shifts are referenced to the internal deuterated solvent resonance and reported as parts-per-million relative to trimethylsilane.

Polymer molecular weights were analyzed *via* gel permeation chromatography (GPC), using a JASCO PU-980 pump JASCO DG-2080-53 degasser, a refractive index detector RI-930, and an UV detector UV-2070. The system was fitted with one guard column PSS SUPREMA 30 Å 10 µm 8x100 mm, one PSS SUPREMA 30 Å 1000 µm 8x300 mm column and one PSS SUPREMA 30 Å 10 µM 8x300 mm column using 10 mM NaN_3 in water with a flow rate of 1.0 mL/min. As standards pullulanes with different molecular weights were used (PSS, Mainz, Germany).

Enzyme assays were performed in BD Falcon non-tissue Culture-treated flat bottom 96-well plates with low evaporation lid (polystyrene) and recorded with a Tecan Infinite M200 Microplates reader.

Isothermal titration calorimetry (ITC) was recorded with a MicroCal VP-ITC calorimeter.

Fluorescence titrations were performed on a JASCO FP-6500 fluorescence spectrometer at 20 °C.

For circular dichroism spectroscopy, a JASCO J-815 spectrometer was used. All measurements were done at 20 °C.

Gelelectrophoreses were performed on a Mini-Protean II from BioRad. A 15 % polyacrylamide gel was used. After electrophoresis the gels were stained with Coomassie-blue.

Capillary zone electrophoresis measurements were performed on a Beckman P/ACE MDQ apparatus with the following separation conditions: fused silica capillary (40 cm effective length, 75 µm internal diameter), hydrodynamic injection for 5 s, temperature 30°C, voltage 20 kV, normal polarity, UV detection at 254 nm. As electrolyte system was used 50 mM citric acid at pH 3.0.

Gel Filtration using JASCO PU-980 pump, JASCO DG-2080-53 degasser, JASCO RI-930 RI detector and UV-2070 UV-detector. The system was fitted with one GE Superdex 200 Increase 10/300 GL column using 100 mM borate buffer with 6 M urea at pH 7.8 with a flow rate of 1.0 mL/min.

Dynamic Light Scattering experiments were performed on a NONO-flex apparatues from Particlemetrix. Boric acid (10 mm) at pH 7.8 was applied as buffer system. An optimal scanning time was found at 40 minutes.

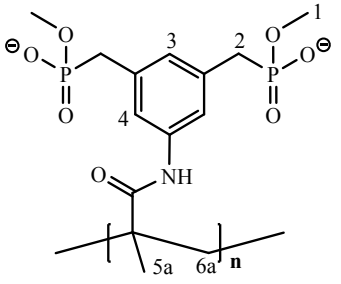
Polymerization Procedures

General Polymerization Procedure
Stock solutions of used monomers were degassed by three freeze-pump-thaw cycles and stored under argon atmosphere prior to use. Under argon atmosphere, the initial volumes of monomer stock solutions, and 5 mol-% AIBN as initiator were then added *via* syringe into a 2 mL vial. The vial was then sealed and placed in a Ditabis HLC-Heating MHL 23 thermomixer at 60 °C and 600 rpm. After 24 h the solutions were lyophilized (Christ, Model Alpha 2-4 LSC).

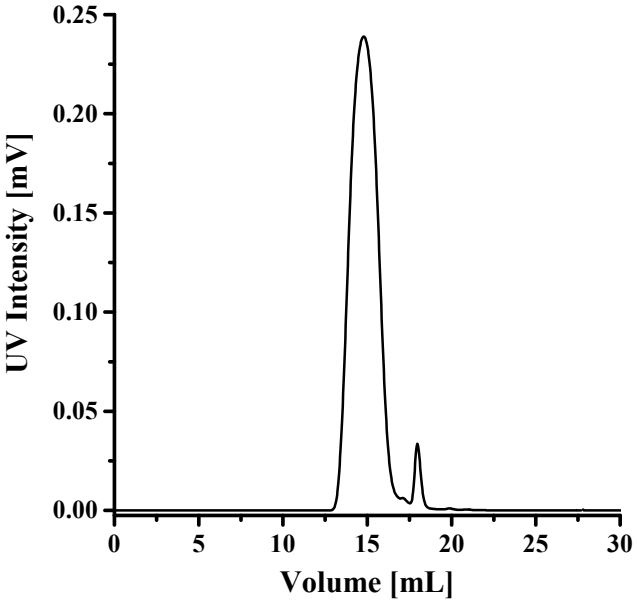
Isolation
The resulting residue was dissolved in 2 mL of ultrapure water and transferred into centrifugal filter unit (Pall Microsep Advance Centrifugal Device 3K MWCO). Ultrafiltration was carried out three times at 4000 rpm for 60 min using an Eppendorf Centrifuge 5706. The resulting oligomer- and monomer-free solution on the filter was subsequently transferred into a 2 mL Eppendorf tube and subjected again to lyophilization.

Characterization
The voluminous colorless lyophilizate was further examined: NMR spectra indicated the degree of conversion and the stoichiometric ratio of comonomers inside the final copolymer. GPC was used for the molecular weight determination of polymers.

P1: Bisphosphonate Homopolymer

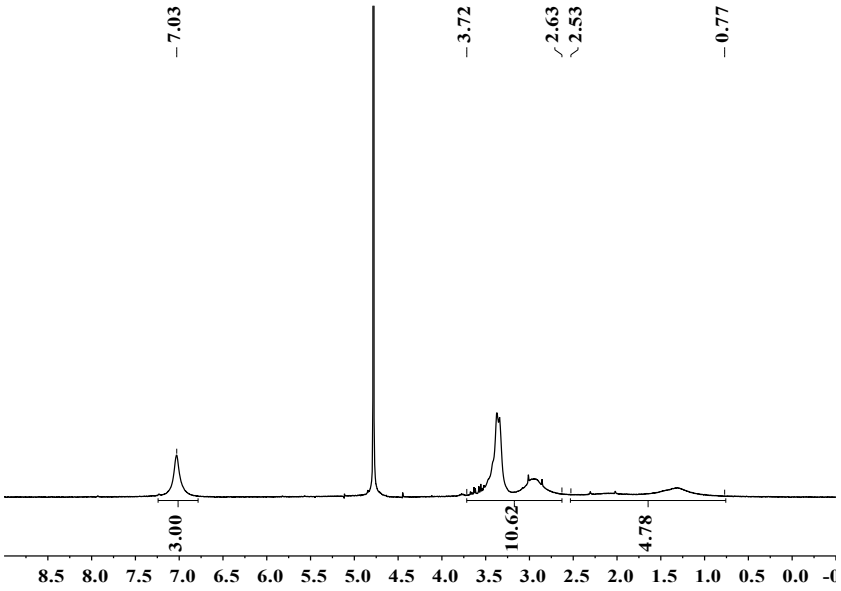


Yield 51.5 mg (62 %).



GPC (Pullulan-Standard):
$\bar{M}_w$ = 236 100 g/mol,
$\bar{M}_n$ = 171 100 g/mol; Đ = 1.38.

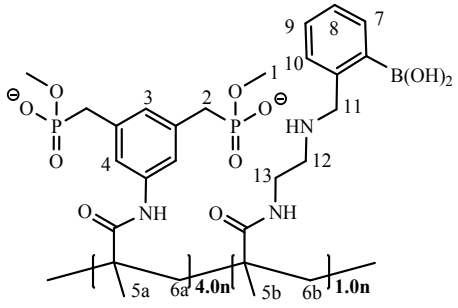
Figure S1. GPC trace of **P1** in 0.01 M NaN_3.



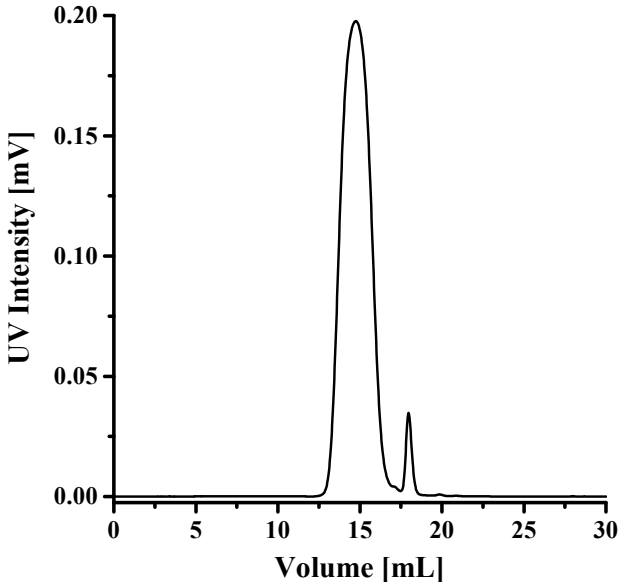
^{1}H NMR (300 MHz, D_2O):
δ [ppm] = **0.77-2.53** (m, 5H, H-5, H-6), **2.63-3.72** (m, 10H, H-1, H-2), **7.03** (sb, 3H, H-3, H-4).

Figure S2. ^{1}H NMR spectrum (300 MHz, D_2O) of **P1**.

P2: Bisphosphonate-Boronate Copolymer



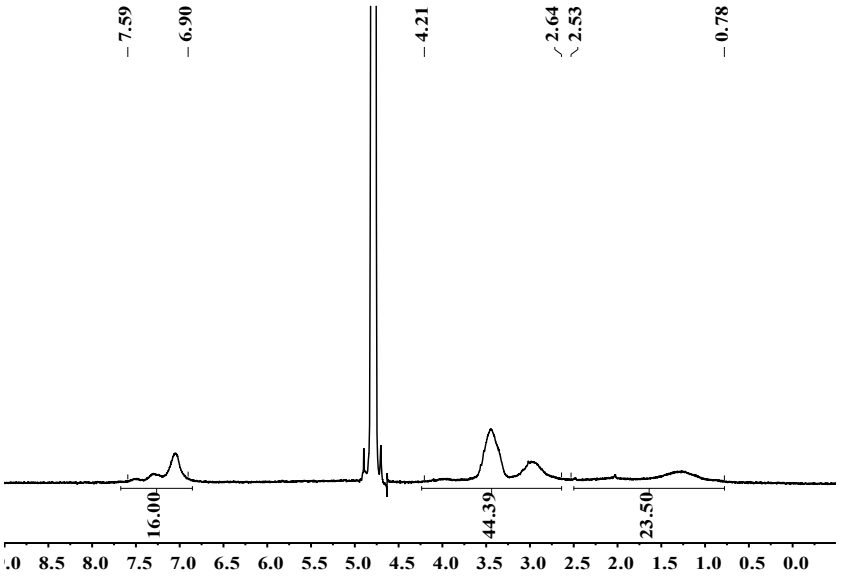
Yield 47.4 mg (55 %).



GPC (Pullulan-Standard):
$\bar{M}_w$ = 264 700 g/mol,
$\bar{M}_n$ = 175 500 g/mol; Đ = 1.51.

Figure S3: GPC trace of P2 in 0.01 M NaN_3.

^{1}H NMR (300 MHz, D_2O):



^{1}H NMR (300 MHz, D_2O): δ [ppm] = **0.78-2.53** (m, 23H, H-5, H-6), **2.61-4.21** (m, 48H, H-1, H-2, H-11, H-12, H-13), **6.90-7.59** (m, 16H, H-3, H-4, H-7, H-8, H-9, H-10).

Figure S4: ^{1}H NMR spectrum (300 MHz, D_2O) of **P2**.

Enzyme Assays

Trypsin Assay

This assay used BAPNA as an artificial substrate.

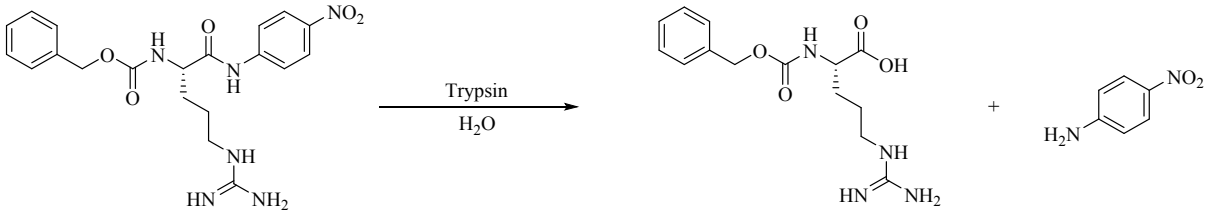


Figure S5. Hydrolytic cleavage of BAPNA with concomitant release of *p*-nitroaniline.

Enzyme and substrate solutions

Preincubation buffer: 100 mM sodium borate, 100 mM HCl, pH 7.8

Trypsin solution: 7.5 µM Trypsin in 1 mM HCl

Substrate solution: 230 µM BAPNA • HCl in H_2O

Incubation buffer: TEA buffer: 200 mM triethanolamine hydrochloride, 20 mM $CaCl_2$, pH 7.8

Procedure:

5 µL enzyme solution were treated with 30 µL borate buffer and 5 µL of polymer solution in 96-well MTP. The mixture was preincubated at 25 °C in darkness for 30 min. Then, 30 µL substrate solution and 70 µL incubation buffer. The MTP was shaken for 10 s and the reaction was measured immediately at 25 °C. Before each scan the MTP was shaken for 4 s. Photometric measurements took place at 405 nm. They were conducted every 30 s for total 30 min.

Stock solutions	Volume	Concentration	Assay concentration
Trypsin	5 µL	7.5 µM	$2.68 \cdot 10^{-7}$ mol/L
Polymer	5 µL	20.0-0.001 mg/mL	$0.71\text{-}3.57 \cdot 10^{-5}$ mg/mL
Preincubation buffer	30 µL	75 mM	
Substrate	30 µL	230 µM	$4.93 \cdot 10^{-5}$ mol/L
Incubation buffer	70 µL	200 mM	
Total volume	140 µL		

Enzyme assay in different buffers

Enzyme and substrate solutions

Preincubation buffer 1: 100 mM potassium dihydrogen phosphate, 100 mM sodium hydroxide, pH 7.8

Preincubation buffer 2: 66.7 mM potassium dihydrogen phosphate, 100 mM sodium phosphate dibasic, pH 7.8

Preincubation buffer 3: 100 mM sodium borate, 100 mM HCl, pH 7.8

Trypsin solution: 7.5 µM Trypsin in 1 mM HCl

Substrate solution: 230 µM BAPNA • HCl in H_2O

Incubation buffer 1: Trypsin buffer TRIS-HCl

Incubation buffer 2: 66.7 mM potassium dihydrogen phosphate, 100 mM sodium phosphate dibasic, pH 7.8

Incubation buffer 3: 100 mM sodium borate, 100 mM HCl, pH 7.8

Incubation buffer 4: TEA buffer: 200 mM triethanolamine hydrochloride, 20 mM $CaCl_2$, pH 7.8

Procedure:

5 µL enzyme solution were treated with 30 µL borate buffer and 5 µL of polymer solution in 96-well MTP. The mixture was preincubated at 25 °C in darkness for 30 min to 120 min. Then, 30 µL substrate solution and 70 µL incubation buffer. The MTP was shaken for 10 s and the reaction was measured immediately at 25 °C. Before each scan the MTP was shaken for 4 s. Photometric measurements took place at 405 nm. They were conducted every 30 s for total 30 min.

Stock solutions	Volume	Concentration	Assay concentration
Trypsin	5 µL	7.5 µM	$2.68 \cdot 10^{-7}$ mol/L
Polymer	5 µL	20.0-0.001 mg/mL	$0.71\text{-}3.57 \cdot 10^{-5}$ mg/mL
Preincubation buffer	30 µL	75 mM	
Substrate	30 µL	230 µM	$4.93 \cdot 10^{-5}$ mol/L
Incubation buffer	70 µL	200 mM	
Total volume	140 µL		

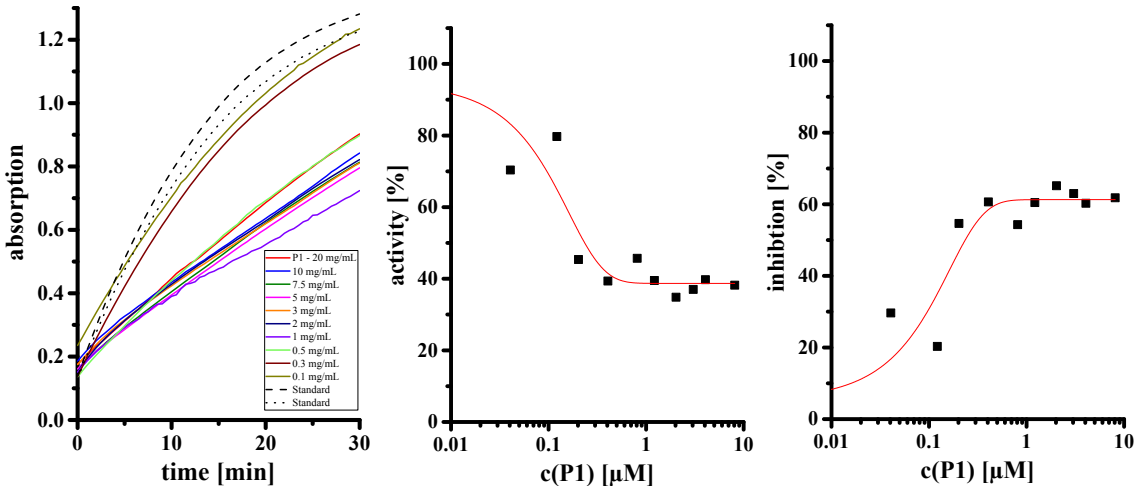


Figure S6. Trypsin Assay with **P1** in Preincubation Buffer 2 (phosphate) after 30 min preincubation at 25 °C in darkness. IC_{50} value 0.235 µM.

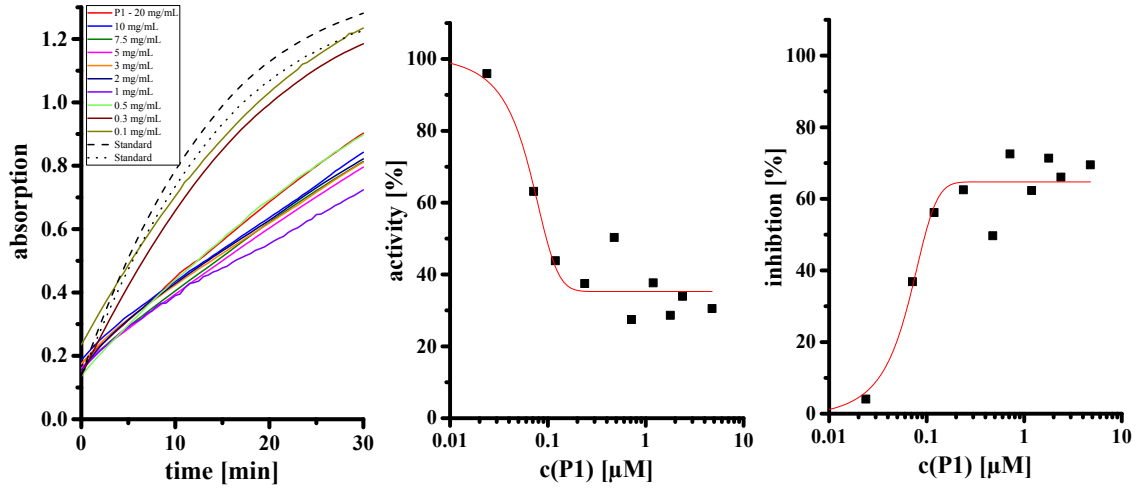


Figure S7. Trypsin Assay with ***P1*** in Preincubation Buffer 2 (phospahte) after 60 min preincubation at 25 °C in darkness. IC_{50} value 0.096 µM.*Figure S8. Trypsin Assay with P1 in Preincubation buffer 2 (phosphate) after 120 min preincubation at 25 °C in darkness. IC_{50} value 0.088 µM.*

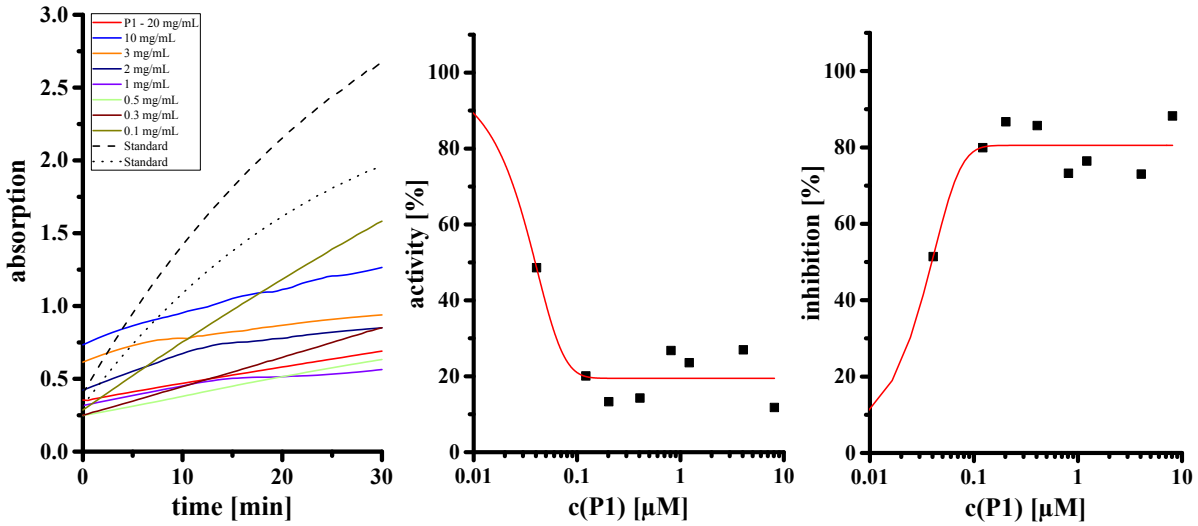


Figure S8. Trypsin Assay with ***P1*** in Preincubation buffer 3 (borate) for 30 min at 25 °C in darkness. IC_{50} value 0.039 µM

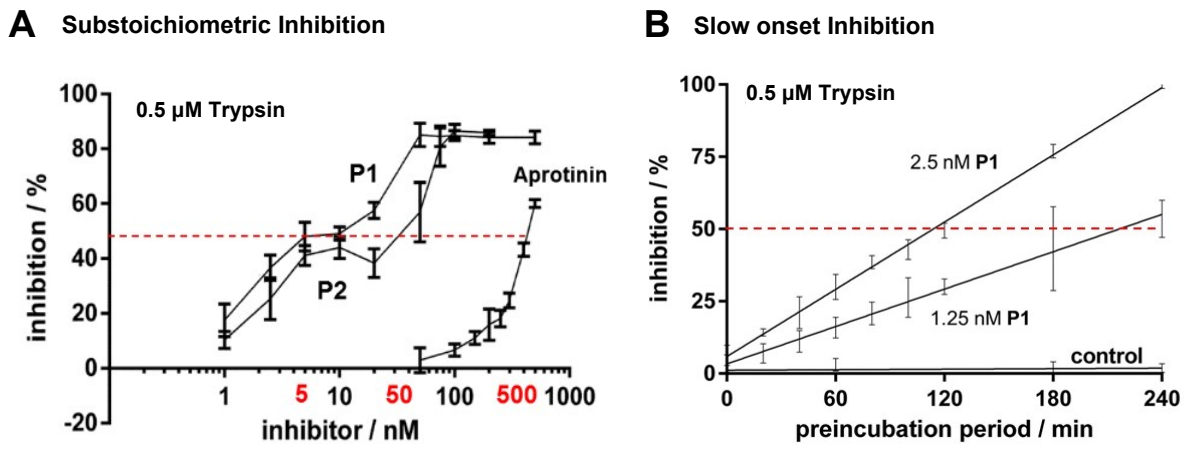


Figure S9. A) Substoichiometric inhibition of trypsin. Affinity polymers demonstrate very low IC_{50} values determined for 500 nM enzyme: P2 50 nM, P1 5 nM. (Each molecule P2 inhibits 10 trypsin molecules, each molecule P1 even 100 enzyme molecules). **B) Slow onset inhibition**. Preincubation of enzyme and polymer further increase inhibition efficiency. After 3h, P1 reaches the IC_{50} at 1 nM.

Reactivation experiment

Enzyme and substrate solutions

Preincubation buffer: 100 mM sodium borate, 100 mM HCl, pH 7.8

Trypsin solution: 28 µM Trypsin in 1 mM HCl

Substrate solution: 230 µM BAPNA • HCl in H_2O

Reactivation solution: 58 µM PEI in H_2O

Incubation buffer: borate buffer: 200 mM triethanolamine hydrochloride, 20 mM $CaCl_2$, pH 7.8

Procedure:

5 µL enzyme solution were treated with 30 µL borate buffer and 5 µL of polymer solution in 96-well MTP. The mixture was preincubated at 25 °C in darkness for 30 min. Then, 30 µL substrate solution and 70 µL incubation buffer. The MTP was shaken for 10 s and the reaction was measured immediately at 25 °C. Before each scan the MTP was shaken for 4 s. Photometric measurements took place at 405 nm. After 10 min 10 µL of Reactivation solution were added and the MTP was shaken for 10 s and hen measured. They were conducted every 30 s for total 40 min.

Stock solutions	Volume	Concentration	Assay concentration
Trypsin	5 µL	28 µM	$2.68 \cdot 10^{-7}$ mol/L
Polymer	5 µL	20.0 mg/mL	$1 \cdot 10^{-6}$ mol/L
Preincubation buffer	30 µL	75 mM	
Substrate	30 µL	230 µM	$4.93 \cdot 10^{-5}$ mol/L
Incubation buffer	70 µL	200 mM	
Reactivation solution	5 µL	58 µM	$2 \cdot 10^{-6}$ mol/L
Total volume	145 µL		

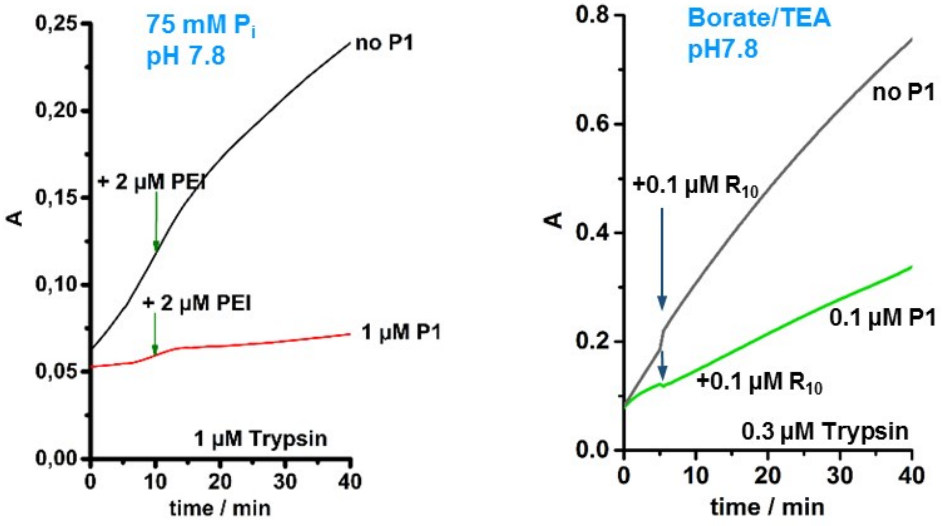


Figure S10. Enzyme assay with and without **P1** after addition of 2 µM PEI (left) or R_{10} (right) to running assay.

Fluorescence Titrations

Stock solutions

Incubation buffer: 100 mM sodium borate, 100 mM HCl, pH 7.8

Polymer solution: 460 nM in incubation buffer

Trypsin solution: 92 mM in polymer solution

Procedure

The polymer stock solution was used to prepare the enzyme stock solution to keep the polymer concentration during titration constant. Then, 700 µL of polymer stock solution were added into a fluorescence cuvette at 20 °C and distinct volumes of the enzyme stock solution were added and well mixed prior to measurement. The excitation wavelength was 330 nm, and the change of fluorescence at 535 nm was observed. A Job plot was used to identify the complex stoichiometry, and for the binding constant a non-linear regression was used.

Polymer	**Kd / nM**
P1	545

Trypsin was titrated to polymer **P1**; the corresponding binding isotherm was evaluated after determination of the complex stoichiometry by a Job plot. The 9:1 protein/polymer ratio indicates that 9 trypsin molecules can be accomodated on the polymer template. Nonlinear regression was subsequently performed for a putative 1:1 complex between one trypsine molecule and a nineth part of the polymer (see Figure S9).

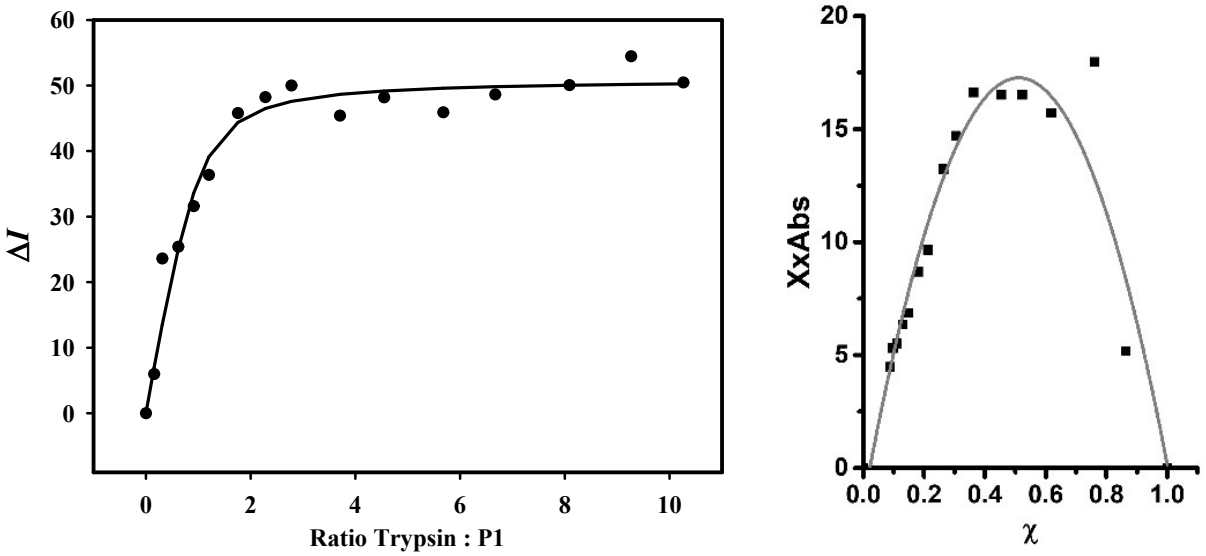


Figure S11. Fluorescence titration of Trypsin with **P1** in phosphate buffer at pH 7.8 and 20 °C. Right: Corresponding Job-Plot normalized to a 1:1 stoichiometry at a 9:1 trypsin/polymer ratio; binding corresponds to a 9:1 stoichiometry.

Isothermal Titration Calorimetry

Stock solutions

Buffer: 75 mM TRIZMA-Base, 75 mM $NaHPO_4$, pH 8.0

Polymer solution: 2.15 µM in buffer

Based on monomer, with respect to DP of 450: 0.96 mM

Trypsinogen solution: 0.1 mM in buffer

	K_a [M^{-1}]	n	ΔH [kcal/mol]	TΔS [kcal/mol]	ΔG [kcal/mol]
P1	n.d.	n.d.	n.d.	n.d.	n.d.
P2	n.d.	n.d.	n.d.	n.d.	n.d.

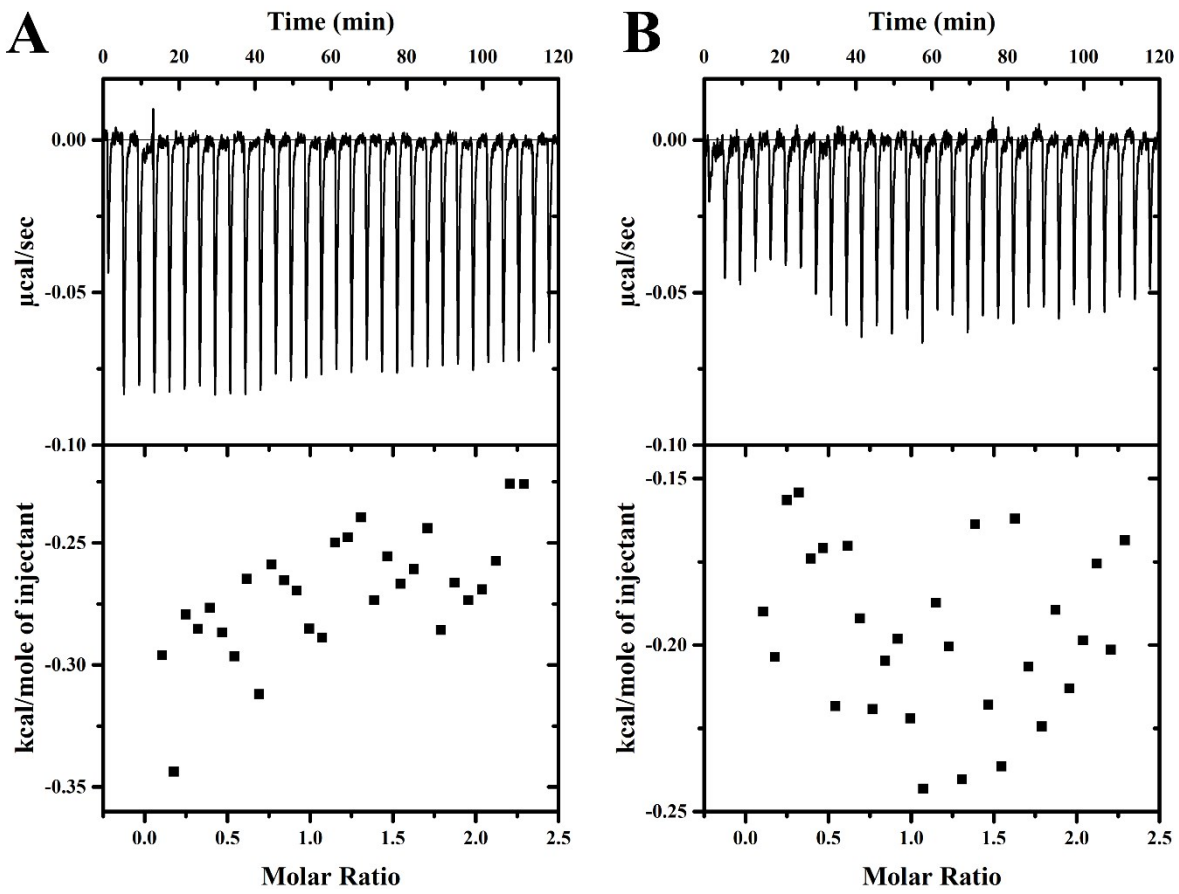


Figure S12. A) Isothermal Titration calorimetry of **P1** and Trypsinogen. B) **P2** vs. Trypsinogen. In both titrations no binding curve can be obtained.

Circular dichroism spectroscopy

Stock solutions

Incubation buffer: 100 mM sodium borate, 100 mM HCl, pH 7.8

Trypsin solution: 0.5 mg in 350 µL incubation buffer, 60 µM

Polymer solution: 67.0 µM in H_2O

Procedure

Incubation of 60 µM Trypsin and 6.7 µM polymer solution in Eppendorf Protein LoBind Tubes at 37 °C and 400 rpm in an Eppendorf Thermomixer Compact. At distinct time points aliquots of 24 µL were taken and diluted in 1976 µL borate buffer in a cuvette. This mixture was measured immediately at 20 °C in the CD spectrometer.

The raw data were smoothed using the *Savitzky-Golay* method (15 pt), 2nd polynomial order.

Incubation		
Solutions	**Volume**	**Concentration**
Trypsin	350 µL	60 µM
Polymer	35 µL	6.7 µM
In cuvette		
Incubation solution		
Trypsin	24 µL	720 nM
Polymer		80 nM
Incubation buffer	1976 µL	100 mM
Total volume	2000 µL	

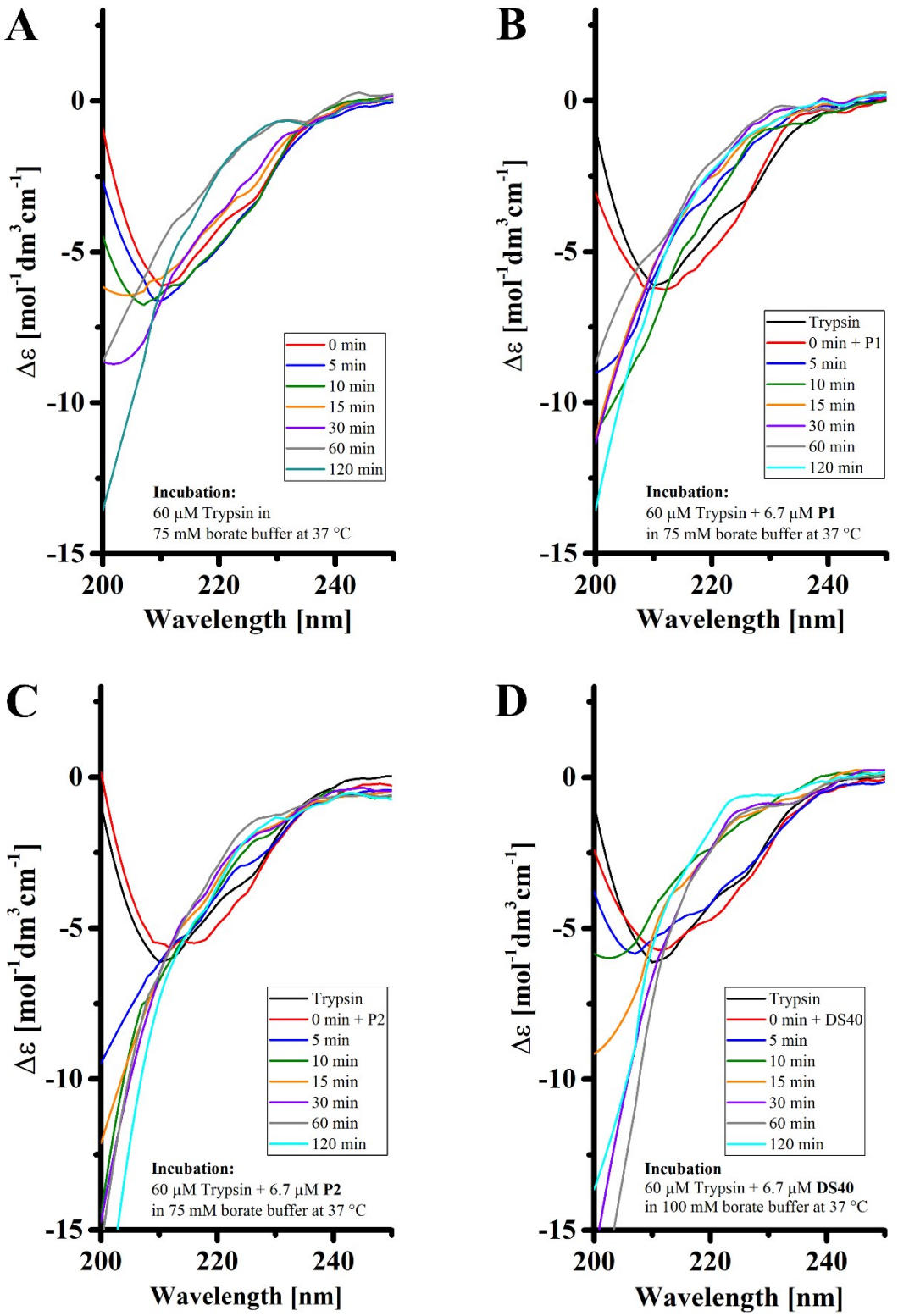


Figure S13. Circular dichroism spectrums of Trypsin in 100 mM borate buffer at 37 °C. A) Autolysis of Trypsin; B) Autolysis of Trypsin with P1; C) Autolysis of Trypsin with P2; D) Autolysis of Trypsin with DS.

Dynamic Light Scattering

Dynamic Light Scattering was carried out on a NONO-flex instrument from Particlemetrix.

Procedure

Experiments were conducted with trypsinogen 250 µM with either **P1** or **P2** (10 µM each) in boric acid (10 mM) at pH 7.8. Four independent experiments were performed at room temperature and one characteristic set is outlined. Unfortunately, a relatively high concentration of trypsinogen had to be applied in order to detect it with certainty. This high concentration yielded artificial aggregates in the presence of **P1** but not in the presence of **P2**.

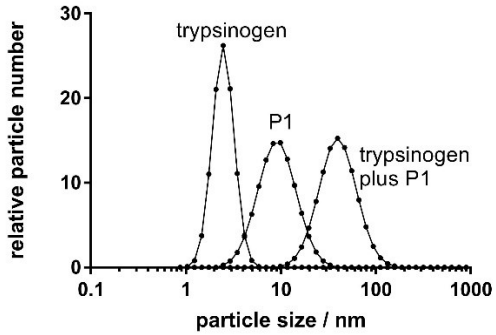


Figure S14. DLS signature of trypsinogen, **P1** and their complex. Note that in this case, a large complex evolves, most likely by formation of unspecific aggregates. This may be a concentration phenomenon, since PFG-NMR testifies compaction to an overall radius of ~7 nm similar to **P2**.

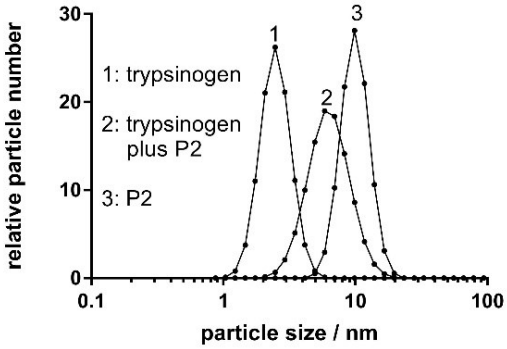
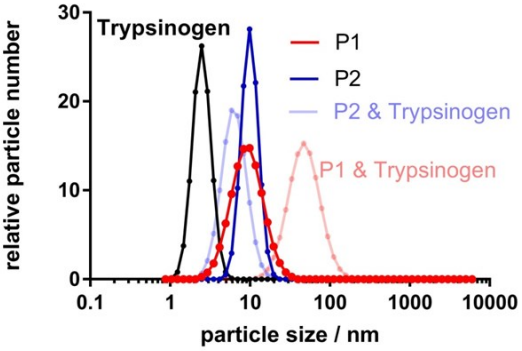


Figure S15. DLS signature of trypsinogen, **P2** and their complex. Note the compaction of the polymer on complex formation to ~ 70% of the original particle size.



Superimposition of Figs. **S14** and **S15**.

Gel Filtration

Stock solutions

Trypsin solution:	2.0 mg Trypsin in 1400 µL incubation buffer, 60 µM
Polymer solution:	67 µM in H_2O
Incubation buffer:	100 mM sodium borate, 100 mM HCl, pH 7.8
Quenching buffer:	6 M urea, 100 mM sodium borate, 100 mM HCl, pH 7.8

Procedure

Incubation of 60 µM Trypsin and 6.7 µM polymer solution in Eppendorf Protein LoBind Tubes at 37 °C and 400 rpm in an Eppendorf Thermomixer Compact. At distinct time points aliquots of 100 µL were taken and mixed with 100 µL cooled quenching buffer and 100 µL 30 mM DTT in water. This mixture was heated 5 min to 95 °C prior to measurement. Then, 200 µL of this mixture were injected to the JASCO HPLC system.

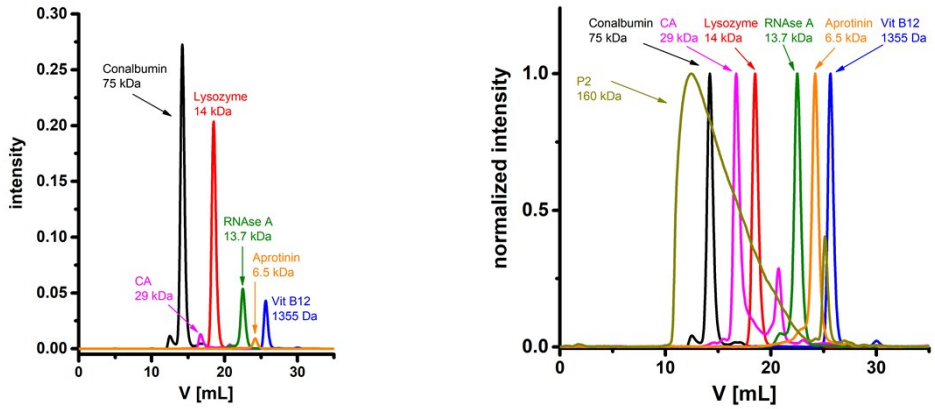


Figure S16. Reference proteins used as standards for calibration of gel filtration measurements. Note that the polymer **P2** is much larger and shows a characteristic tailing.

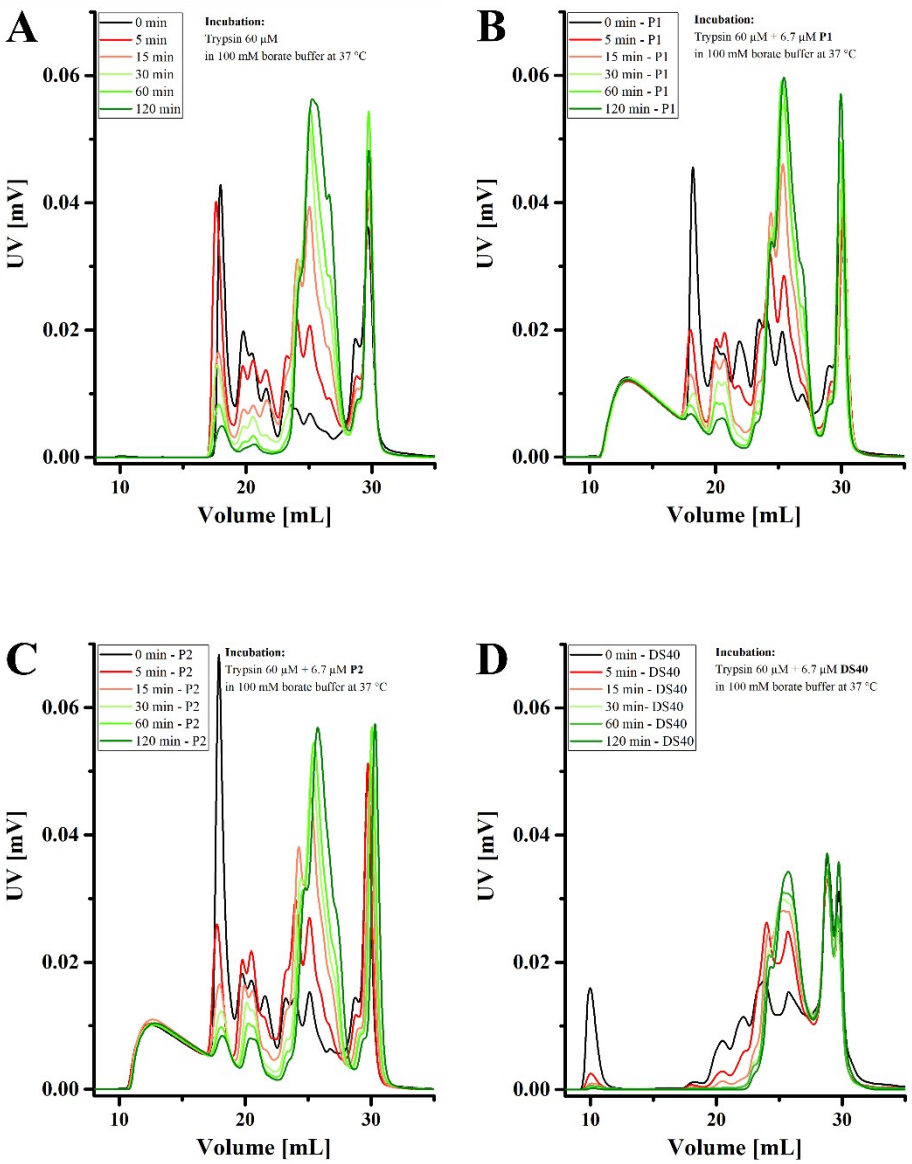


Figure S17. Elugrams of Trypsin autolysis in 100 mM borate buffer. A) autolysis of Trypsin, B) autolysis of Trypsin with **P1**, C) autolysis of Trypsin with **P2**, D) autolysis of Trypsin with DS40. Note the drastic decrease of the main band (β-trypsin) after 5 min induced by the presence of **P1** and **P2**.

Capillary Zone Electrophoresis

Comment [h]:

Procedure

Capillary Zone Electrophoresis was carried out on a Beckman P/ACE MDQ instrument. Experiments monitored the disappearence of native trypsin (50/75 µM). Tryptic self-digest without additive, monitored at different time intervals, revealed that even after 70 min, more than 35% of intact trypsin is still present. By contrast, addition of 0.5 mol-% **P1** or of 0.8 mol-% **P2** completely eliminated the trypsin band after 17 min.

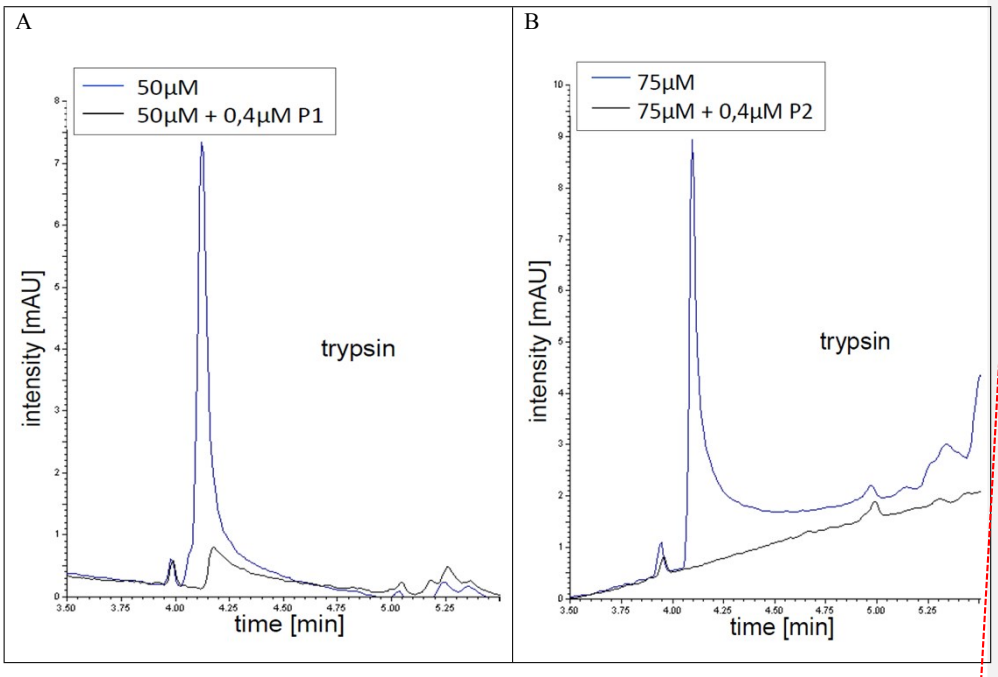


Figure S18. Chromatograms of trypsin in the absence and in the presence of **P1** or **P2** (0.4 µm each) after a reaction time of 8 min at 25°C in 10 mm borate buffer (pH 7.8). A) 50 µm trypsin; B) 75 µm trypsin.

SDS-PAGE

Stock solutions

Buffer A:	1.5 M Tris-HCl, pH 8.8, 0.4 % SDS (w/v)
Buffer B:	0.5 M Tris-HCl, pH 6.8, 0.4 % SDS (w/v)
Loading buffer:	0.3 M Tris-HCl, 10 % SDS (w/v), 40 % glycerol, pH 6.8, 0.001 % bromophenol blue, 30 mM DTT
Running Buffer:	333 mM Tris-HCl, 1.92 M glycine, 1% SDS (w/v), pH 8.3
Trypsin stock:	1.1 mM in 1 mM HCl, stored on ice
Compound stocks:	67 µM in Tris buffer
Solution A:	25 % Isopropanol, 10 % acetic acid, 65 % water (v/v), 0.05 % Coomassie R250 (w/v)
Solution D:	10 % acetic acid, 90 % water (v/v)

Procedure

Prewarming 152.2 µL buffer to 37 °C. addition of 9.82 µL Trypsin solution and 18 µL of compound stock solution. Incubation of his mixtures at 37 °C and 400 rpm. At distinct time points aliquots of 12 µL were taken and quenched in 10 µL of Tris buffer with 30 mM DTT. These probes were cooled in ice, then heating to 90 °C for 5 min, then centrifuged for 1 min. From this solution 10 µL were loaded on 15 % Polyacrylamide gel. Gelelectrophoresis was done in SDS running buffer for 30 min with a constant current of 200 V.

The gels were stained with Coomassie-blue following the procedure[9] heating in 600 W microwave for 1 min in solution A, followed by 5 min gentle shaking. Solution D was used for destaining of background, first heating 1 min in microwave, then at RT for 2 h.

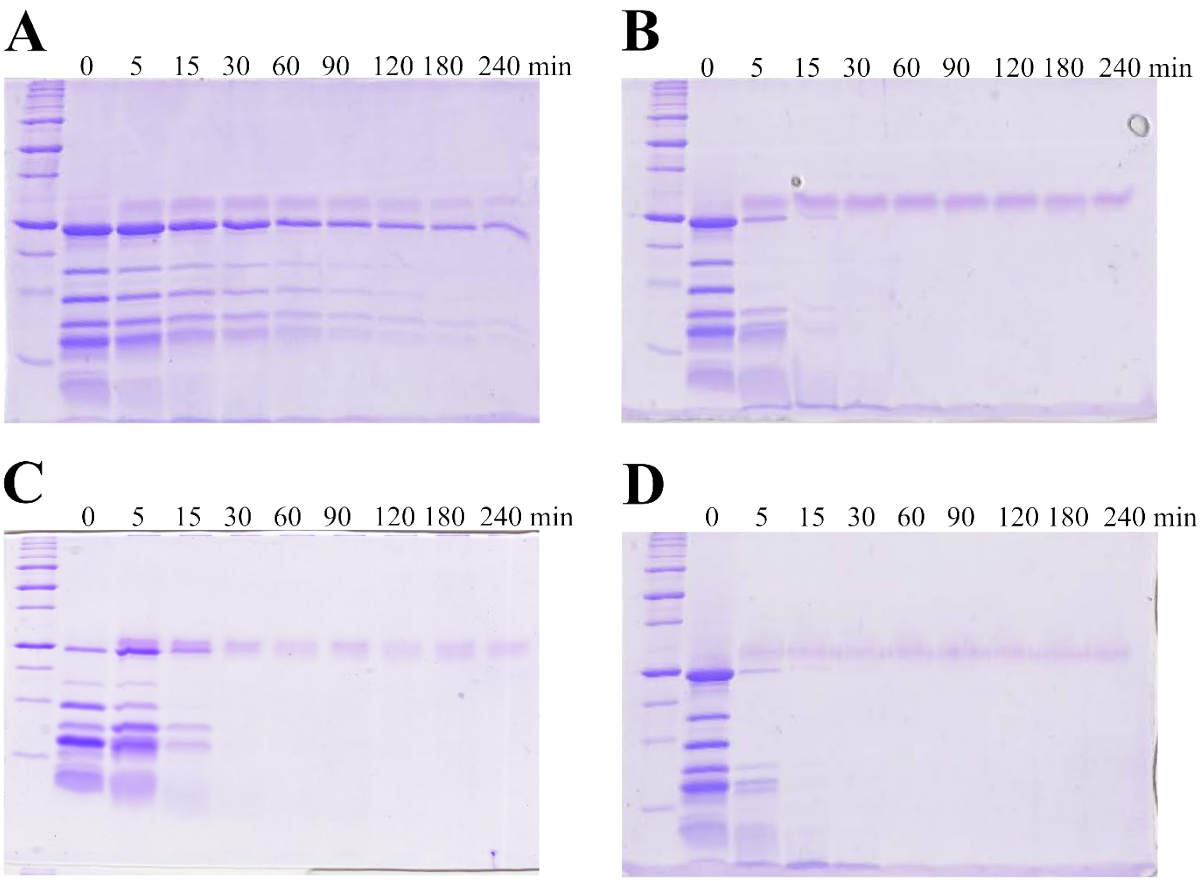


Figure S19. Autolysis of Trypsin in 75 mM TRIS buffer. A) Trypsin; B) 60 µM Trypsin and 6.7 µM **P1**; C) 60 µM Trypsin and 6.7 µM **P2**; D) 60 µM Trypsin and 6.7 µM DS40.

Pulsed-Field Gradient Nuclear Magnetic Resonance (PFG-NMR)

^{1}H NMR diffusion experiments were run on a 400 MHz Bruker Avance spectrometer with a Bruker DIFF30 probe head. All measurements were performed at 298 K.

Both polymer and protein were dissolved in borate buffer (100 mM) at pH = 7.8. The borate buffer was prepared with D_2O to obtain a lock signal in the NMR experiment. The water peak is also decreased by D_2O. To suppress the water peak, the PFG-NMR experiment was additionally performed with a presaturation experiment.[1]

For sample preparation, 200 µL polymer (559 µmol) and 200 µL protein (520 µmol) were filled in regular 5 mm NMR tubes. For all measurements, the stimulated echo pulse sequence with two gradient pulses was used. Sixteen scans were accumulated for each setting. The time between two gradient pulses Δ was 50 ms. The gradients were adjusted to strengths G between 30 and 500 G/cm with a duration δ of 1.5 ms. All measurements (the full set of gradient strengths under the variation from 30 to 500 G/cm) were repeated four times.

Principle: PFG-NMR can be used to study the molecular diffusion in the sample solution. The PFG-NMR technique combines the ability to reveal information on the chemical nature as well as on the molecular or collective translational mobility of the individual components. The observed molecules, may be assigned to a structural part of the dispersion depending on its characteristic motional behaviour.[1,2] In a solution, the free self-diffusion coefficients of the dissolved molecules can be determined with the PFG-NMR. In the given case, all PFG-NMR experiments consist of the application of two field gradient pulses with a stimulated echo pulse sequence (90°−τ1−90°−τ2−90°−τ1-echo). The pulse gradients with a gradient strength G and duration δ are applied during both of the waiting periods τ1 with an overall separation Δ. In the presence of free diffusion with a diffusion coefficient D, this leads to a decay of the echo intensity I with respect to the original value I_0 (for G = 0) according to

$$I/I_0 = I_{rel} = exp[-\gamma^2\delta^2G^2D(\Delta - \delta/3)] \qquad (1)$$

The negative value of the apparent diffusion constant can be calculated from the slope in the Stejskal-Tanner plot (ln I_{rel} versus $\gamma^2\delta^2G^2(\Delta - \delta/3)$).[4-6]

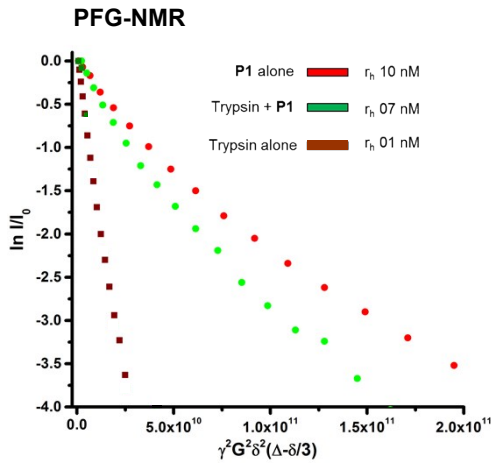


Figure S20. Stejskal−Tanner plot[1] obtained from the PFG-NMR measurements. The slope of the straight line gives the negative value of the diffusion coefficient of the investigated molecule. From this the hydrodynamic polymer, protein and complex radii were calculated, confirming DLS results (trypsin 1.3 nm, polymer 10.3 nm, complex 7.1 nm).

References PFG-NMR Spectroscopy

[1] Altieri, AS and Byrd, RA. Randomization approach to water suppression in multidimensional NMR using pulsed field gradients. Journal of Magnetic Resonance, Series B 107(3):260--266, 1995

[2] J., Linders; C., Mayer; T., Sekine; H., Hoffmann: Pulsed-Field Gradient NMR Measurements on Hydrogels from Phosphocholine. In: J. Phys. Chem. B 2012 116, S. 11459-11465.

[3] D., Molnar; J., Linders; C., Mayer; R., Schubert: Insertion stability of poly(ethylene glycol)-cholesteryl-based lipid anchors in liposome membranes. In: European Journal of Pharmaceutics and Biopharmaceutics, 2016, 103, S. 51-61

[4] E. O. Stejskal, J. E. Tanner: Spin Diffusion Measurements. Spin Echoes in the Presence of a Time-Dependent Field Gradient. In: The Journal of Chemical Physics, 1965, 42 (1), S. 288-292.

[5] E. O., Stejskal: Use of Spin Echoes in a Pulsed Magnetic-Field Gradient to Study Anisotropic, Restricted Diffusion and Flow. In: The Journal of Chemical Physics, 1965, 43 (10), S. 3597-3603.

[6] J. E., Tanner; E. O., Stejskal: Restricted Self-Diffusion of Protons in Colloidal Systems by the Pulsed-Gradient, Spin-Echo Method. In: The Journal of Chemical Physics, 1968, 49 (4), S. 1768-1777.

Visualization of the Self-Digest Mechanism at elevated concentrations

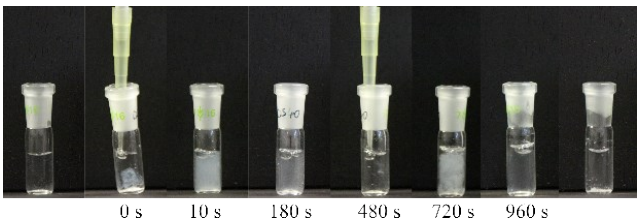


Figure S21. Photographic snapshots of the polymer-catalyzed trypsin self-digest, monitoring the progress and illustrating the typical phases of polymer-assisted self-digest: 1. formation of the polymer/protein complex with low net charge (precipitate); 2. templated proteolysis with concomitant product dissociation from the polymer (gradually clearing solution); 3. addition of more protein

[1] E. O. Stejskal, J. E. Tanner: Spin Diffusion Measurements. Spin Echoes in the Presence of a Time-Dependent Field Gradient. *The Journal of Chemical Physics*, **1965**, *42*, 288-292.

substrate (new precipitate); 4. templated proteolysis with concomitant product dissociation from the polymer (gradually clearing solution).

Mass Spectrometry

Time course experiment for mass spectrometry
In order to analyse the effect of polymer P1 on the activity of trypsin a time course experiment was carried out (Fig. S19). A trypsin stock solution (60 µM) in borate buffer (100 mM; pH 7.8) was prepared at 0°C and treated with or without Polymer P1 (67 µM based on average molecular mass of 170000 g/mol). After removal of the t=0 samples (triplicates) and immediate precipitation by adding 5 volumes of ice cold 5% formic acid (FA) in acetone, the stock was transferred to 37°C and incubated for 5, 10, 15, 30 and 60 min. At the indicated time points samples were taken, precipitated with 5 volumes of ice cold 5% FA in acetone and then centrifuged (18000 ×g, 5 min, 4°C). 5% FA stops trypsin activity; Acetone precipitates undigested trypsin and P1 while most of the trypsin digestion products are soluble in Acetone; the subsequent centrifugation pellets the higher molecular weight components. The supernatant after centrifugation was transferred to a fresh Eppendorf tube and the organic solvent removed in a vacuum concentrator. The dried pellets, containing the trypsin digestion products, were then cleaned-up for LC-MS.

Sample clean-up for LC-MS. Acidified tryptic digests were desalted on home-made C18 StageTips as described[1]. On each 2 disc StageTip we loaded around 15 µg peptides (based on the initial protein concentration). After elution from the StageTips, samples were dried using a vacuum concentrator (Eppendorf) and the peptides were taken up in 10 µL 0.1 % formic acid solution.

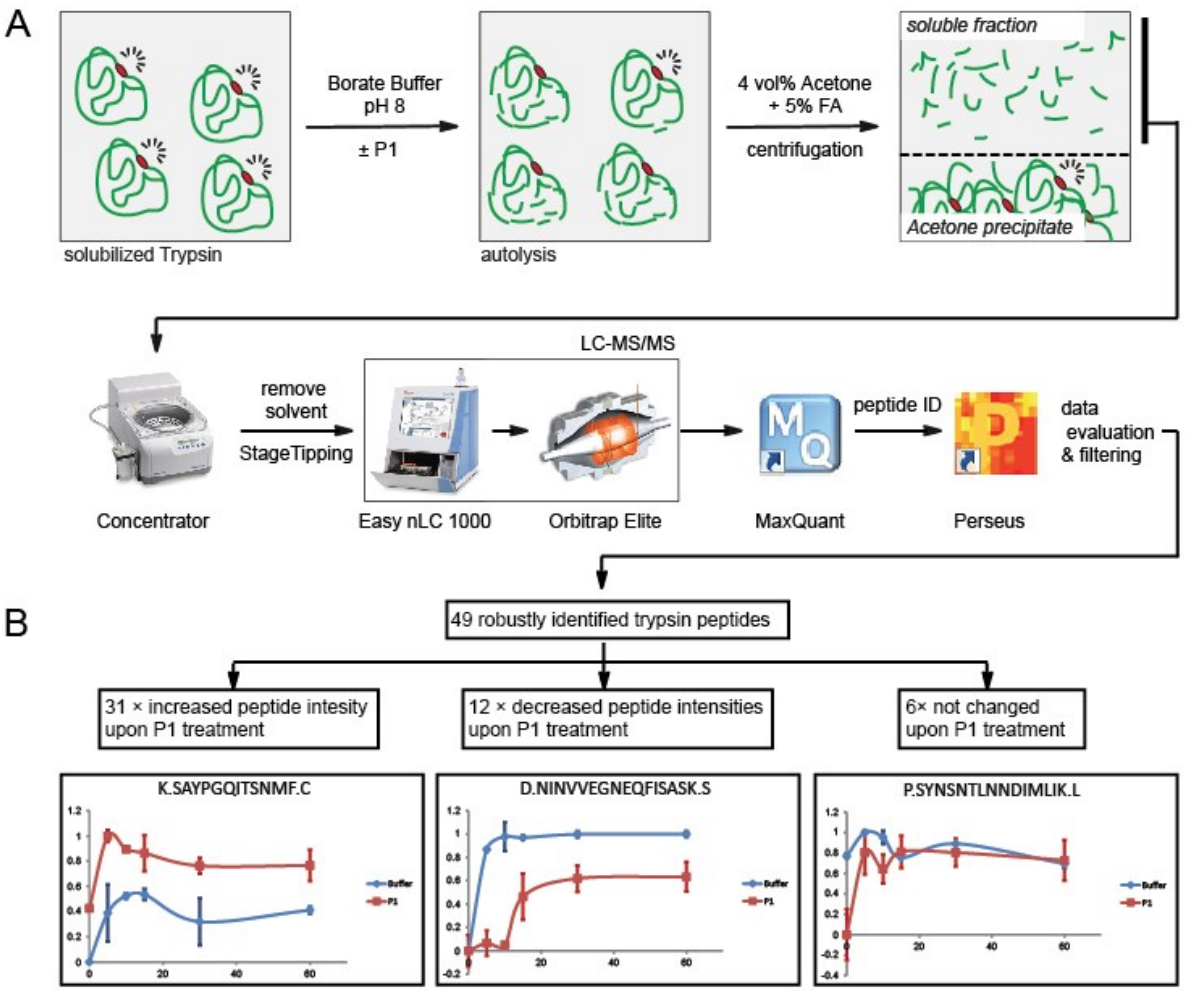
LC-MS/MS
Experiments were performed on an Orbitrap Elite instrument (Thermo)[2] coupled to an EASY-nLC 1000 liquid chromatography (LC) system (Thermo) operated in the one-column mode. The analytical column was a fused silica capillary (75 µm × 22 cm) with an integrated PicoFrit emitter (New Objective) packed in-house with Reprosil-Pur 120 C18-AQ 1.9 µm resin (Dr. Maisch). The analytical column was encased by a column oven (Sonation) and attached to a nanospray flex ion source (Thermo). The column oven temperature was adjusted to 45 °C. The LC was equipped with two mobile phases: solvent A (0.1% formic acid, FA, in water) and solvent B (0.1% FA in acetonitrile, ACN). All solvents were of UPLC grade (Sigma). Peptides were directly loaded onto the analytical column with a flow rate around 0.5 – 0.8 µL/min, which did not exceed 980 bar. Peptides were subsequently separated on the analytical column by running a 70 min gradient of solvent A and solvent B (start with 7% B; gradient 7% to 35% B for 60 min; gradient 35% to 80% B for 5 min and 80% B for 5 min) at a flow rate of 300 nl/min. The mass spectrometer was set in the positive ion mode and operated using Xcalibur software (version 2.2 SP1.48). Precursor ion scanning was performed in the Orbitrap analyzer (FTMS; Fourier Transform Mass Spectrometry) in the scan range of m/z 300-1500 and at a resolution of 60000 with the internal lock mass option turned on (lock mass was 445.120025 m/z, polysiloxane)[3]. Product ion spectra were also recorded in a data-dependent fashion in the FTMS in a variable scan range. The ionization potential was set to 1.8 kV. Peptides were analyzed using a repeating cycle consisting of a full precursor ion scan (1.0×10^6 ions or 50 ms) followed by 12 product ion scans (1.0×10^4 ions or 100 ms) where peptides are isolated based on their intensity in the full survey scan (threshold of 500 counts) for tandem mass spectrum (MS2) generation that permits peptide sequencing and identification. Higher-energy collisional dissociation (HCD) normalized collision energy was set to 30% for the generation of MS2 spectra. During MS2 data acquisition dynamic ion exclusion was set to 120 seconds with a maximum list of excluded ions consisting of 500 members and a repeat count of one. Ion injection time prediction, preview mode for the FTMS, monoisotopic precursor selection and charge state screening were enabled. Only charge states higher than 1 were considered for fragmentation.

Peptide and Protein Identification using MaxQuant
RAW spectra were submitted to an Andromeda[4] search in MaxQuant (version 1.5.3.30) using the default settings.[5] Label-free quantification and match-between-runs was activated.[6] MS/MS spectra data were searched against the *Bos taurus* (UP000009136_9913.fasta; 24150 entries, downloaded 6/16/2017). To estimate the level of contamination, all searches included a contaminants database (as

implemented in MaxQuant, 245 sequences) that contains known MS contaminants. Andromeda searches allowed oxidation of methionine residues (16 Da) and acetylation of the protein N-terminus (42 Da) as dynamic modifications and no static modifications. Digestion mode was set to "semispecific", Enzyme specificity was set to "Trypsin/P" with 2 missed cleavages allowed, the instrument type in Andromeda searches was set to Orbitrap and the precursor mass tolerance to ±20 ppm (first search) and ±4.5 ppm (main search). The MS/MS match tolerance was set to ±0.5 Da and the peptide spectrum match FDR and the protein FDR to 0.01 (based on target-decoy approach and decoy mode "revert"). Minimum peptide length was 7 amino acids. Minimum score for unmodified peptides was set to 0. All peptide relevant evidence data can be found in supplemental Table S20. For protein quantification modified peptides (minimum score 40) and unique and razor peptides were allowed. Retention times were recalibrated based on the built-in nonlinear time-rescaling algorithm. MS/MS identifications were transferred between LC-MS/MS runs with the "Match between runs" option in which the maximal match time window was set to 0.7 min and the alignment time window set to 20 min. The quantification is based on the "value at maximum" of the extracted ion current. [6] Further analysis and annotation of identified peptides was done in Perseus v1.5.5.3.[7] Processed data can be found in supplemental Tables S20. For the analysis we only used peptides mapping to *B. taurus* trypsin (Uniprot Nr. P00760). Only peptides that were identified in at least two independent MS runs and with a peptide score above 80 were considered for further analysis. For quantification related technical replicates were combined to categorical groups and normalized within one MS run by subtraction of the "median". Then the intensity average (based on "median") for each categorical group was calculated together with the standard deviation (SD). The averaged intensities were then filtered again removing all peptides with less than 6 valid values. The normalized, averaged and scaled peptide intensities were then used to plot the peptide profiles over time (Supplemental File S20). Peptides were then sorted according to their profiles in three categories: 1) increased abundance upon P1 treatment, 2) increased abundance without P1 and 3) no effect of P1. The C- and N-terminal amino acid of each category 1 Peptide was mapped onto the X-Ray structure of P00760 (PDB: 1bty). Sites in red are the C-terminal amino acids of the detected peptide sequences (cut after); the N-terminal amino acids are depicted in blue (cut before). Amino acids in yellow depict peptides that were found at the C- and N-terminus of different peptides (either cut before or after).

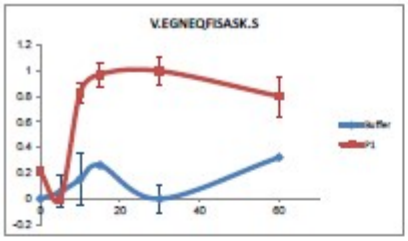
Data availability

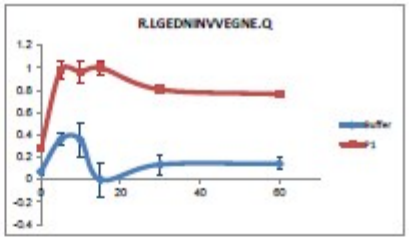
The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [8] partner repository (https://www.ebi.ac.uk/pride/archive/) with the dataset identifier PXD017969. During the review process the data can be accessed via a reviewer account (Username: reviewer38133@ebi.ac.uk; Password: yK5FsXjw).

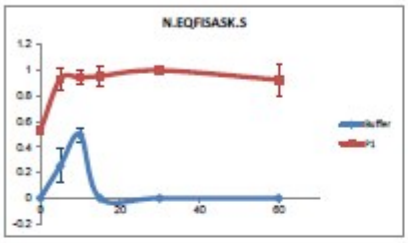


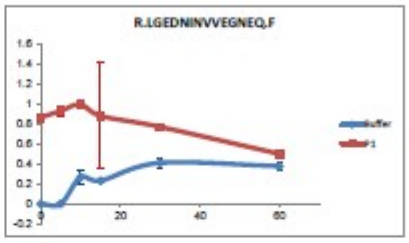
Supplementary Figure S22. Affinity Polymers P1/P2 cleave with low specificity at uncomplexed surface areas. A) Terminating Trypsin self-digest with formic acid and isolation of soluble peptide cleavage products after precipitation of large protein material by acetone. **B)** Representative kinetic diagrams after LC-MS/MS proteomics analysis, exhibiting the effects of **P1**. Left: accelerated cleavage (31x), center: decreased proteolysis (6x), and c) unaltered cleavage kinetics (12x).

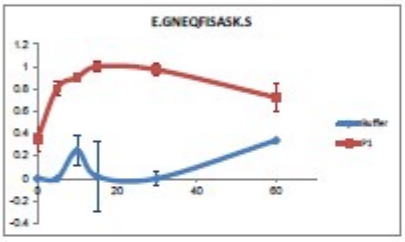
Peptide Intensity increased upon P1 treatment (1 of 3)

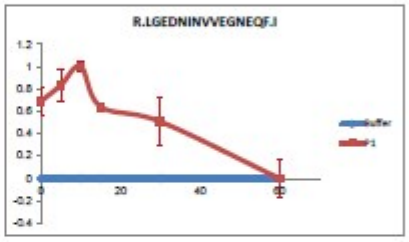


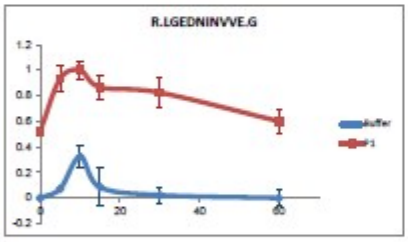


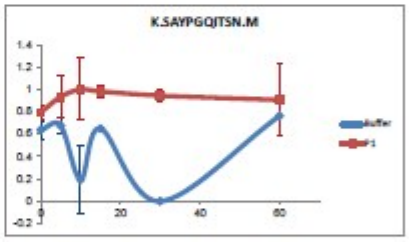


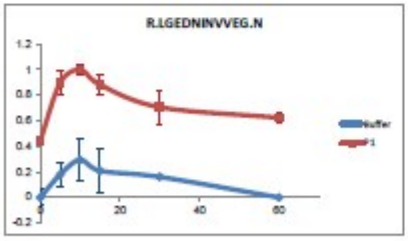


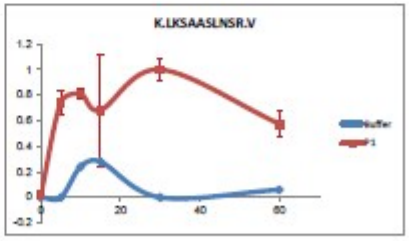


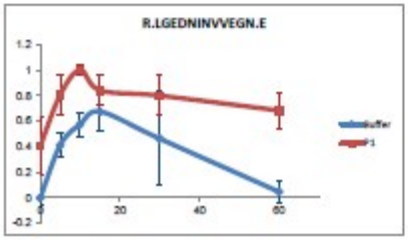


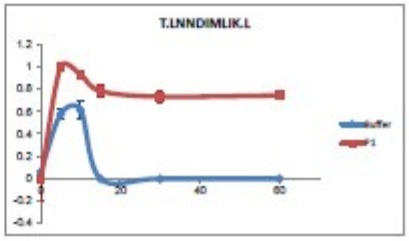




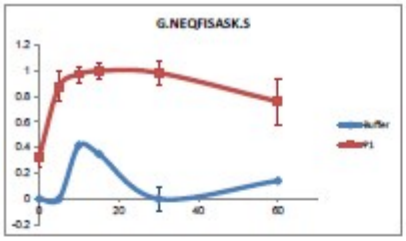


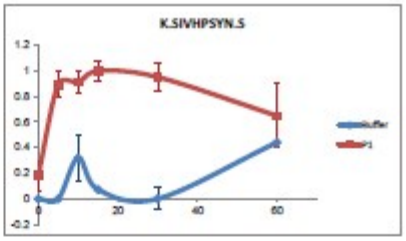


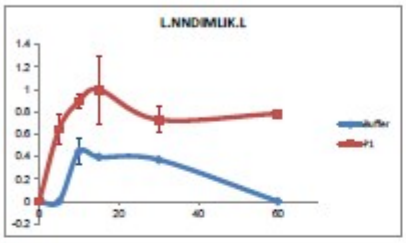


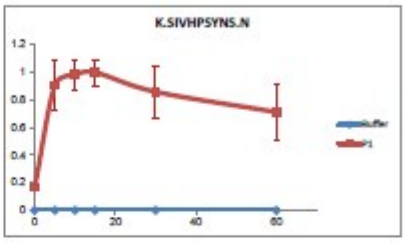


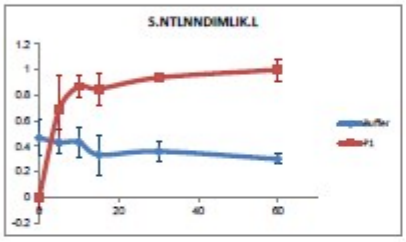
Peptide Intensity increased upon P1 treatment (2 of 3)

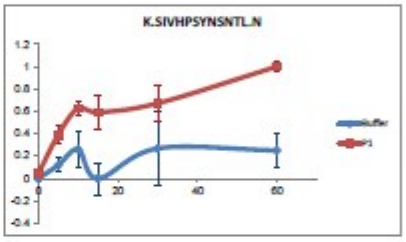


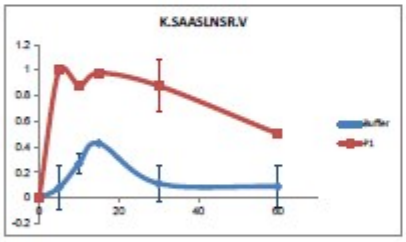


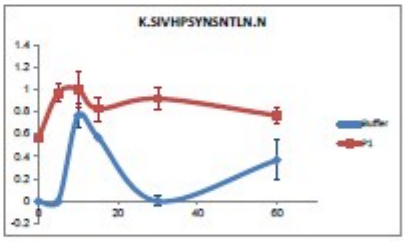


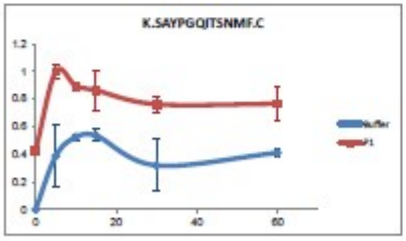


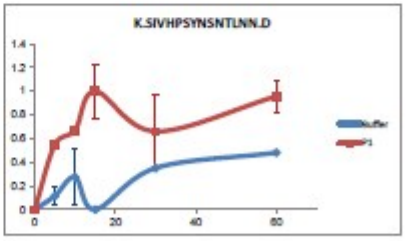


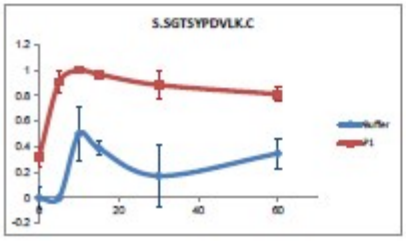


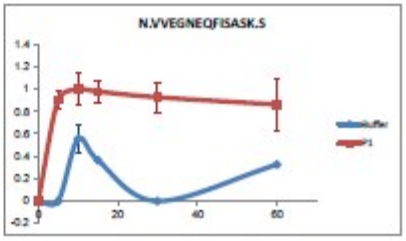




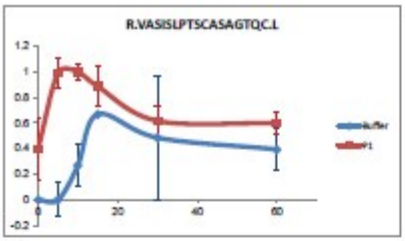


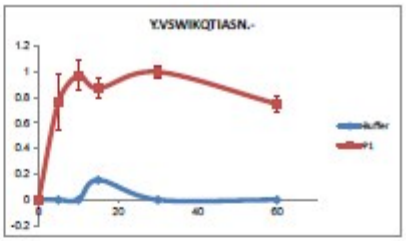


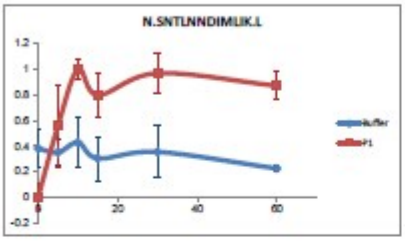


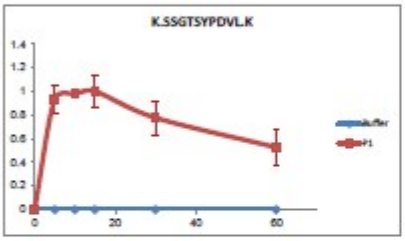


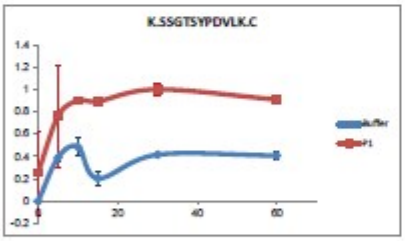
Peptide Intensity increased upon P1 treatment (3 of 3)

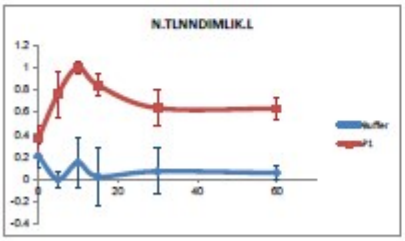


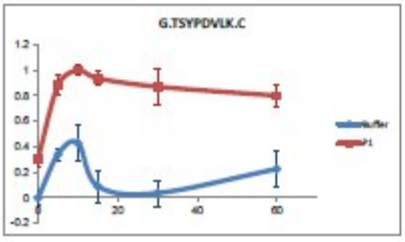






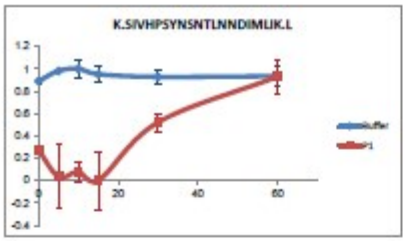


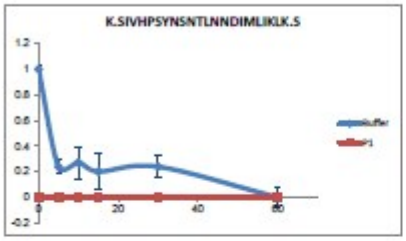


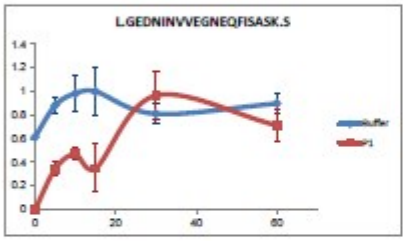


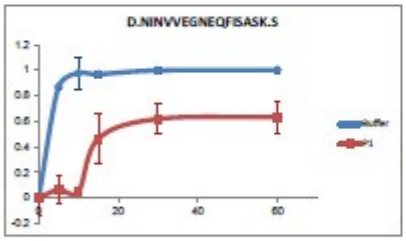
31 peptides were produced with strongly accelerated cleavage kinetics compared to native trypsin alone.

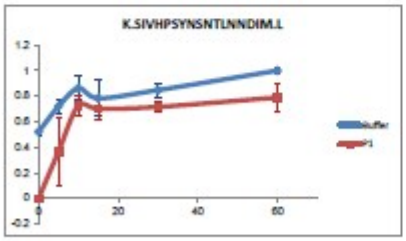
Peptide Intensity decreased upon P1 treatment

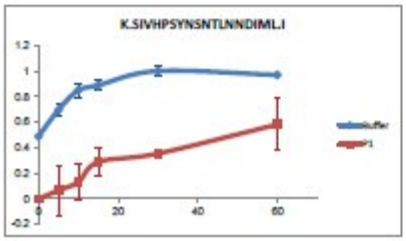








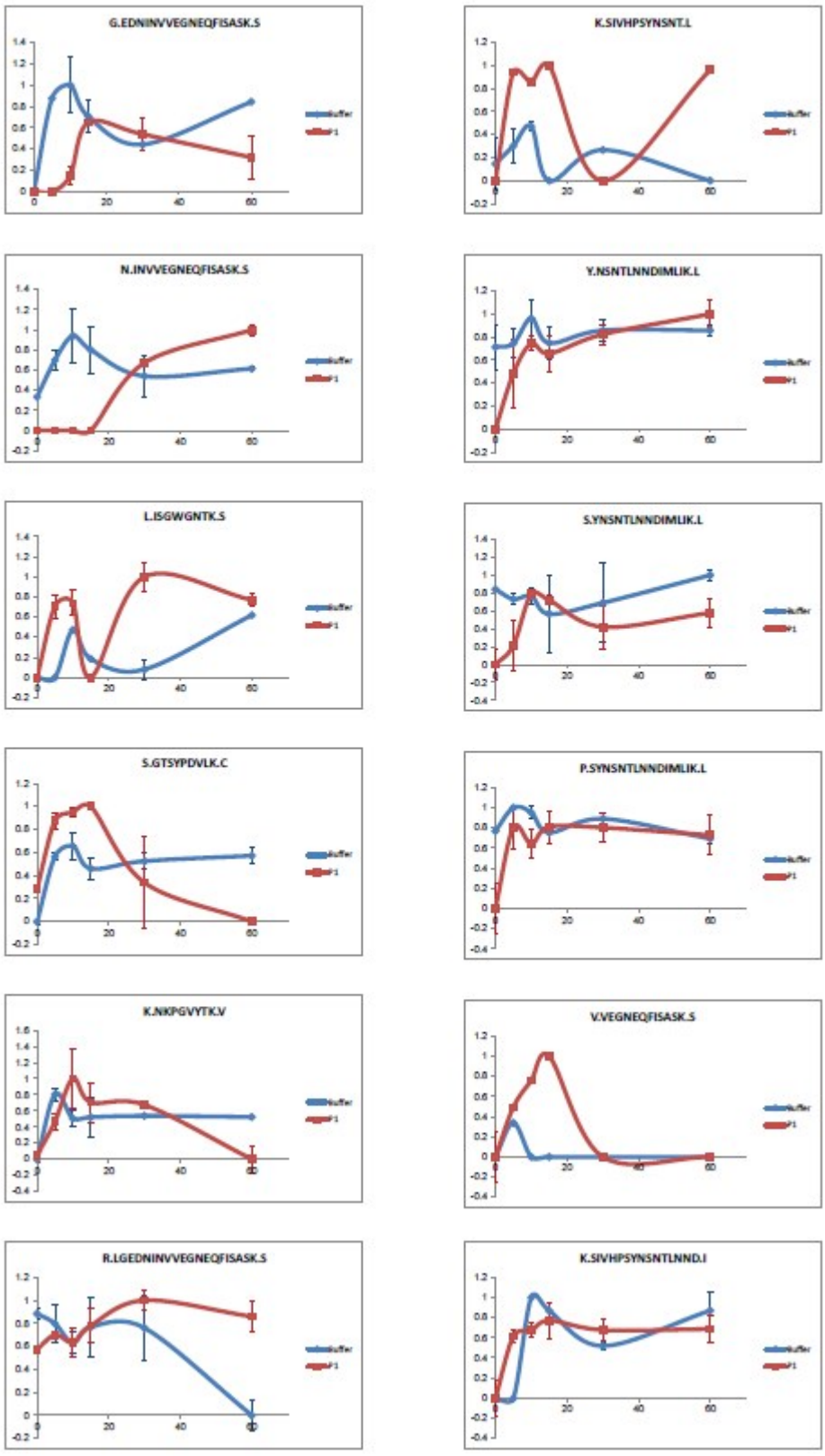




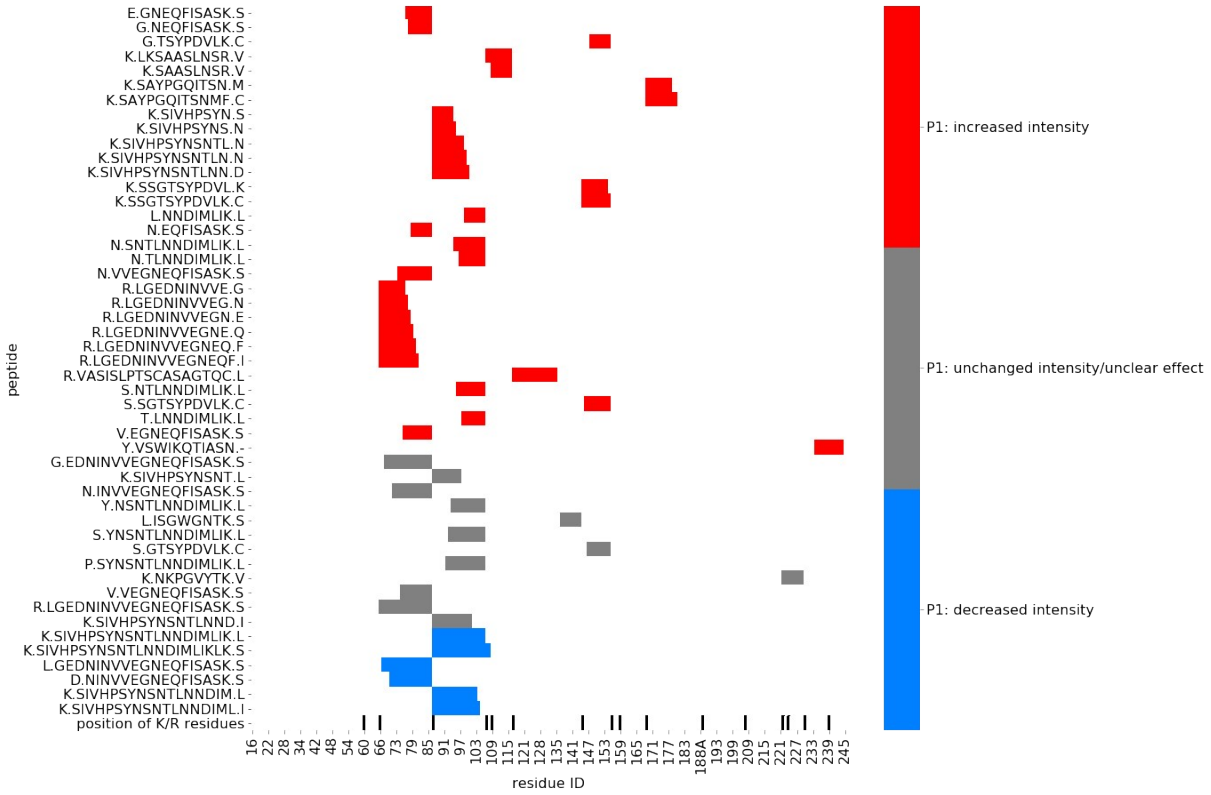
6 peptides were produced with significantly decelerated cleavage kinetics during autolysis of native trypsin.

S31

no or unclear effect of P1



12 peptides were produced without significantly altered cleavage kinetics during autolysis of native trypsin.



Supplementary Figure S23. Peptide Sequences aligned to Trypsin primary sequence. Peptides in red show increased, peptides in grey unchanged and peptides in blue decreased abundance upon **P1** treatment.

Many cleavages from detected peptides are unlikely to be initial proteolytic events. Some cleavage sites are buried within the protease or located within secondary structure elements, i.e. they cannot be cleaved before other cleavage and/or major denaturation has taken place (Figure S24).

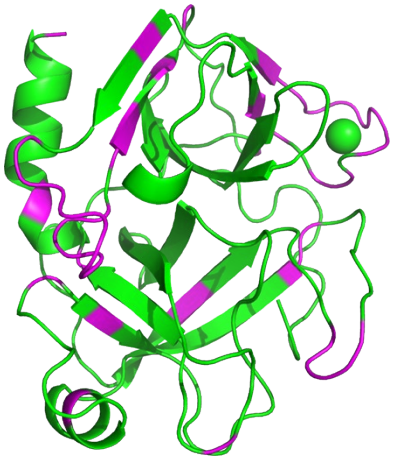


Figure S24. Cleavage sites from all detected peptides. All amino acids located as direct neighbours C-terminally to the cleaved peptide bond are colored in magenta.

References Mass Spectrometry

[1] J. Rappsilber, M. Mann, Y. Ishihama, *Nat. Protoc.* **2007**, *2*, 1896-1906.

[2] A. Michalski, E. Damoc, O. Lange, E. Denisov, D. Nolting, M. Muller, R. Viner, J. Schwartz, P. Remes, M. Belford, J. J. Dunyach, J. Cox, S. Horning, M. Mann, A. Makarov, *Mol. Cell. Proteomics* **2012**, *11*, O111 013698.

[3] J. V. Olsen, L. M. de Godoy, G. Li, B. Macek, P. Mortensen, R. Pesch, A. Makarov, O. Lange, S. Horning, M. Mann, *Mol. Cell. Proteomics* **2005**, *4*, 2010-2021.

[4] J. Cox, N. Neuhauser, A. Michalski, R. A. Scheltema, J. V. Olsen, M. Mann, *J. Proteome Res.* **2011**, *10*, 1794-1805.

[5] J. Cox, M. Mann, *Nat. Biotechnol.* **2008**, *26*, 1367-1372.

[6] J. Cox, M. Y. Hein, C. A. Luber, I. Paron, N. Nagaraj, M. Mann, *Mol. Cell. Proteomics* **2014**, *13*, 2513-2526.

[7] S. Tyanova, T. Temu, P. Sinitcyn, A. Carlson, M. Y. Hein, T. Geiger, M. Mann, J. Cox, *Nat. Methods* **2016**, *13*, 731-740.

[8] J. A. Vizcaino, A. Csordas, N. Del-Toro, J. A. Dianes, J. Griss, I. Lavidas, G. Mayer, Y. Perez-Riverol, F. Reisinger, T. Ternent, Q. W. Xu, R. Wang, H. Hermjakob, *Nucleic Acids Res.* **2016**, *44*, 11033.

[9] C. Wong, S. Sridhara, J. C. Bardwell, U. Jakob, *Biotechniques* **2000**, *28*, 426-428, 430, 432.

Computational Methods

Lattice model

To model interactions of full polymers with multiple trypsin molecules we considered a regular cubic lattice in a cubic simulation box containing one polymer molecule surrounded by N protease molecules (Figure 3E). In the simulations presented here, the box had a length of 32 lattice distances and a volume of $32^3 = 32768$ lattice positions. Each particle – i.e. a protease molecule or a monomer of the polymer molecule – occupies a single position on the lattice. Each particle is potentially multivalent for binding to particles of the other kind; specifically, a single monomer can bind a maximum of four (termini: five) protease molecules, and a protease can bind up to six monomers of the polymer . There is only one energy parameter: we assume that each protease-monomer contact (protease and monomer at neighboring lattice positions) is rewarded with the same attractive energy term, E_{cont}. Particles are not allowed to overlap. Accordingly, polymer conformations are modeled as self-avoiding walks on the cubic lattice.

For a given simulation box containing a single polymer molecule, our computational model has three independent parameters: DP, the degree of polymerization of the polymer molecule; N, the number of proteases in the simulation box; E_{cont}, the affinity (in units of k_BT) per contact of a trypsin molecule with a monomer of the polymer. For our computational simulations we chose parameter values that approximately reflected the experimental situation. Since DP in the experiment covered a range of values, we simulated for several DP values, namely 10, 20, and 30. $DP = 30$ is probably shorter than the longest polymers in the experiment but the covered range should reveal trends while still being computationally tractable. The number $N = 10$ trypsin molecules per polymer molecule was chosen to mirror the experimental ratio of concentrations of about 10 protease molecules per polymer molecule. For the contact energy we assumed values of $E_{cont} = -2k_BT$ and $-4k_BT$ (corresponding to about 5 to $10kJ/mol$ at $T = 300K$) that are typical of weak to medium affinities in biomolecular systems.

Thermodynamic averages of quantities such as polymer size or number of protease-protease contacts were computed with the following algorithm. We performed a series of n lattice Monte Carlo simulations where we first produced a large number n of random polymer conformations $i = 1,\ldots,n$ as self-avoiding random walks of length DP on the cubic lattice (Landau and Binder 2009). Rosenbluth weights w_i were computed for all walks (Rosenbluth and Rosenbluth 1955). For each random walk i, N protease molecules were introduced at random free lattice positions, and we then sampled a large number of trypsin arrangements around the fixed polymer conformation i by random moves to any lattice position using the Metropolis criterion (Metropolis et al. 1953). Thermodynamic averages were then computed in two steps. First, we averaged quantities of interest x for each polymer conformation i to obtain averages $\langle x \rangle_i$, e.g. the average energy $\langle E \rangle_i$. Second, from these averages a global thermodynamic average was evaluated as $\langle x \rangle = \sum_{i=1}^{n} \langle x \rangle_i$ with probabilities $p_i = q_i \Big/ \sum_{j=1}^{n} q_j$ and weight factors $q_i = w_i \cdot exp\left(-\langle E \rangle_i / (k_BT)\right)$, where k_B is the Boltzmann constant and T the temperature. The thermodynamic averages R_m, n_{ee}, n_{ep}, and n_b shown in Figure 3D were computed with this two-step procedure. Details on these quantities are given in the following.

The mean monomer-monomer distance R_m of a polymer with given DP was evaluated by first sampling $n_{fp} = 4 \cdot 10^5$ conformations j of the free polymer (without proteases) and the corresponding Rosenbluth weights w_j. From this polymer-only simulation we estimated a reference value $R_{m,DP}^{(ref)}$ for the given DP as

$$R_{m,DP}^{(ref)} = \sum_{j=1}^{n_{fp}} w_j \cdot \frac{2}{DP \cdot (DP-1)} \cdot \sum_{1 \le k < l}^{DP} |r_{jk} - r_{jl}|$$

with r_{jk} the lattice coordinate vector of monomer k of polymer conformation j. We then evaluated the corresponding values $R_{m,DP}$ in the presence of protease molecules with the Monte Carlo method described above and reported in Figure 3D(a) the values of R_m (with index DP omitted for brevity), defined as ratio $R_m = R_{m,DP}/R_{m,DP}^{(ref)}$. Thus $R_m < 1$ indicates a collapse of the polymer from the protease-free polymer to the polymer in presence of proteases.

To compute the number n_{ee} of protease-protease contacts we monitored the number of protease molecules at neighboring lattice positions, i.e. at positions that differed by 1 lattice unit in either x or y or z direction during Monte Carlo sampling at the end of each sweep of N protease repositioning trials. This was then averaged in the two step procedure as described above to a number n'_{ee}. Similar to the case of R_m we used a reference number of contacts $n_{ee}^{(ref)}$ computed in absence of protease-polymer affinity, and reported $n_{ee} = n'_{ee}/n_{ee}^{(ref)}$ in Figure 3D(b), i.e. the amplification of protease-protease contact frequency by the polymer.

The number n_{ep} of protease-polymer contacts was defined as number of pairs of protease molecules and monomers at positions that differed by 1 lattice unit in either x or y or z. Since each such polymer-protease contact contributes an energy of E_{cont}, the total energy directly informs about the number of protease-polymer contacts. The n_{ep} values in Figure 3D(c) were obtained by the two-step averaging procedure.

The number n_b of protease molecules bound to the polymer in Figure 3D(d) was determined by monitoring counts of the protease molecules with at least one contact to the polymer as defined previously, and the two-step averaging procedure.

To check convergence of thermodynamic averages we monitored changes of their variances between Monte Carlo simulation blocks of increasing lengths (Flyvbjerg and Petersen 1989). Error bars in Figure 3D are differences between 5% and 95% quantiles of the thermodynamic averages if evaluated separately for the first and second half of the data simulated for the respective parameter combination of DP and E_{cont}.

Model and algorithms have been implemented in the Julia programming language (https://julialang.org), and are freely available as source code at https://github.com/DanielHoffmann32/LatticePolymers.jl.

Epitopsy

To compute the affinity map of a bisphosphonate monomer around a trypsin molecule (Figures 3C, 4A-C) we proceeded as follows. The bisphosphonate ligand geometry (Table SBP) was calculated in OpenBabel v2.3.2 (O'Boyle et al. 2011) starting from a SMILES string. Van der Waals radii were added automatically by OpenBabel. Atomic partial charges were calculated by the Gasteiger PEOE (Gasteiger et al. 1980) method in PyBabel v1.4alpha1 via AutoDockTools v1.5.4 (Morris et al. 2009, Sanner 1999) from the MGLTools v1.5.6 suite. The trypsin structure was obtained from PDB 2PTN (Walter 1982). Charges, van der Waals radii and missing hydrogen atoms were added by PDB2PQR v1.9.0 (Dolinsky et al. 2007, Dolinsky et al. 2004) at neutral pH with the Amber force field option. The trypsin electrostatic field was calculated by solving the non-linear Poisson–Boltzmann equation with APBS version 1.4.1 (Baker et al. 2001) with ionic concentration 0.15 mol/L and relative dielectric permittivities $\varepsilon_r^{vacuum} = 2$ and $\varepsilon_r^{water} = 79$. The trypsin environment was scanned with the BP monomer in Epitopsy

(Grad et al. 2018) using 150 rotations and a grid resolution of 0.8 Å. Note that this approach will not reveal effects of polymer multivalency.

Trypsin autolysis is much faster if Ca^{2+} ions are absent. Ca^{2+} is attributed with stabilizing trypsins structure, reducing its susceptibility to proteolysis, rather than increasing its activity (Nord et al. 1951). We note that our calculations of trypsin-BP affinities use a crystal structure (PDB entry 2PTN) including Ca^{2+} because there is no trypsin structure without it in the protein database. The effect of possible structural changes of trypsin due to the lack of Ca^{2+} in our experiments is unclear.

Table SBP: BP monomer conformation used in Epitopsy calculations. Positions and radii are given in Å, charges in *e*.

	X	Y	Z	Charge	Radius
C	5.321	1.170	2.351	0.0360	1.7000
C1A	5.624	2.368	2.998	-0.0510	1.7000
C2A	6.246	2.379	4.254	-0.0440	1.7000
C3A	6.583	1.161	4.857	-0.0700	1.7000
C4A	6.269	-0.062	4.243	-0.0440	1.7000
C5A	5.642	-0.045	2.985	-0.0510	1.7000
H1A	5.407	-0.999	2.516	0.0870	1.1000
H2A	5.387	3.328	2.552	0.0870	1.1000
H3A	7.096	1.171	5.817	0.0850	1.1000
N	4.697	1.109	1.078	-0.3230	1.5500
C0D	4.289	2.178	0.302	0.2510	1.7000
C1D	3.602	1.811	-0.992	-0.0030	1.7000
C2B	6.647	-1.368	4.917	0.0550	1.7000
C2D	3.760	0.638	-1.628	-0.1110	1.7000
C3B	6.531	3.696	4.944	0.0550	1.7000
C3D	2.759	2.908	-1.591	-0.0470	1.7000
C8B	3.183	-2.303	6.699	0.0140	1.7000
C9B	9.115	4.978	2.253	0.0140	1.7000
O	4.379	3.365	0.577	-0.2680	1.5200
O1P	4.637	-3.048	4.074	-0.6460	1.5200
O2P	5.941	-3.622	6.280	-0.6460	1.5200
O3P	4.332	-1.660	6.197	-0.3170	1.5200
O4P	9.274	3.642	4.826	-0.6460	1.5200
O5P	8.055	5.912	5.325	-0.6460	1.5200
O6P	7.933	4.866	3.014	-0.3170	1.5200
P1	5.346	-2.598	5.335	0.2530	1.8000
P2	8.106	4.579	4.609	0.2530	1.8000
H	4.517	0.173	0.752	0.1690	1.1000
H0B	7.387	-1.881	4.291	0.0590	1.1000
H1B	7.150	-1.148	5.866	0.0590	1.1000
H1C	2.583	-2.711	5.881	0.0670	1.1000
H1D	4.399	-0.158	-1.265	0.0670	1.1000

	X	Y	Z	Charge	Radius
H2B	5.718	4.395	4.706	0.0590	1.1000
H2C	3.461	-3.102	7.392	0.0670	1.1000
H2D	3.244	0.446	-2.564	0.0670	1.1000
H3B	6.474	3.545	6.029	0.0590	1.1000
H3C	2.581	-1.568	7.239	0.0670	1.1000
H3D	1.767	2.915	-1.128	0.0330	1.1000
H4C	8.857	4.860	1.197	0.0670	1.1000
H4D	3.208	3.898	-1.457	0.0330	1.1000
H5C	9.836	4.198	2.509	0.0670	1.1000
H5D	2.632	2.775	-2.671	0.0330	1.1000
H6C	9.568	5.963	2.396	0.0670	1.1000

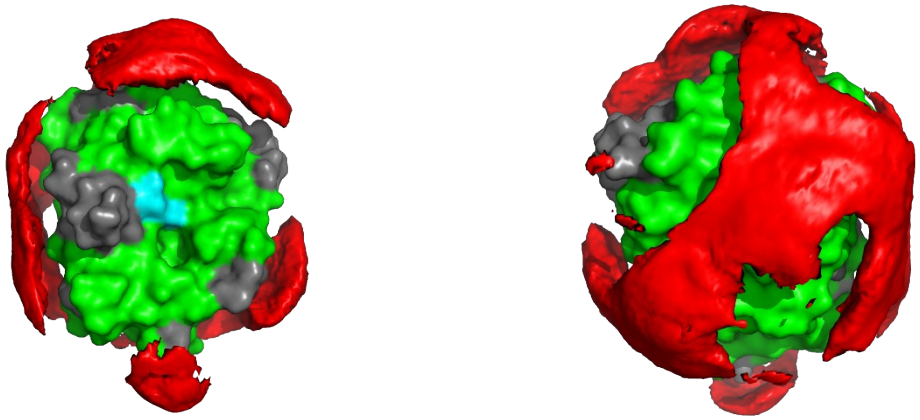


Figure S25. Epitopsy surface for bisphosphonate monomer (red) and cleavage sites on Trypsin in the presence of **P1** (grey). The active site is labeled in blue. Left: Front view; right: back view. Note the complementarity of both areas on the protein surface – accelerated cleavage occurs mainly in the open space.

References Computation

Baker, NA, Sept D, Joseph S, Holst MJ, and McCammon JA. 2001. "Electrostatics of nanosystems: application to microtubules and the ribosome." *Proc Natl Acad Sci U S A* 98: 10037–41.

Dolinsky TJ, Nielsen JE, McCammon JA, and Baker NA. 2004. "PDB2PQR: An Automated Pipeline for the Setup of Poisson-Boltzmann Electrostatics Calculations." *Nucleic Acids Res* 32 (Web Server issue): W665–667. doi:10.1093/nar/gkh381.

Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, and Baker NA. 2007. "PDB2PQR: Expanding and Upgrading Automated Preparation of Biomolecular Structures for Molecular Simulations." *Nucleic Acids Research* 35 (Web Server issue): W522–W525. doi:10.1093/nar/gkm276.

Flyvbjerg H, and Petersen HG. 1989. "Error Estimates on Averages of Correlated Data." *The Journal of Chemical Physics* 91 (1). AIP Publishing: 461–66. doi:10.1063/1.457480.

Gasteiger J, Marsili M. “Iterative partial equalization of orbital electronegativity – a rapid access to atomic charges“. *Tetrahedron*. 1980; 36: 3219–3228. doi:10.1016/0040-4020(80)80168-2

Grad JN, Gigante A, Wilms C, Dybowski JN, Ohl L, Ottmann C, Schmuck C, and Hoffmann D. 2018. “Locating Large, Flexible Ligands on Proteins.” *Journal of Chemical Information and Modeling* 58 (2). American Chemical Society (ACS): 315–27. doi:10.1021/acs.jcim.7b00413.

Landau DP, and Binder K. 2009. *A Guide to Monte Carlo Simulations in Statistical Physics*. Cambridge: Cambridge University Press.

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, and Teller E. 1953. “Equation of State Calculations by Fast Computing Machines.” *Jcp* 21 (6): 1087–92.

Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. “AutoDock4 and autodocktools4: Automated docking with selective receptor flexibility“. *Journal of Computational Chemistry*. 2009; 30: 2785–2791. doi:10.1002/jcc.21256

O’Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. „Open babel: An open chemical toolbox“ *Journal of Cheminformatics.* 2011; 3: 33. doi:10.1186/1758-2946-3-33

Rosenbluth MN, and Rosenbluth AW. 1955. “Monte Carlo Calculation of the Average Extension of Molecular Chains.” *The Journal of Chemical Physics* 23 (2). AIP Publishing: 356–59. doi:10.1063/1.1741967.

Sanner MF. “Python: A programming language for software integration and development“. *Journal ofMolecular Graphics & Modelling*. 1999; 17: 57–61.

Schrödinger, LLC. 2015. “The PyMOL Molecular Graphics System, Version 1.8.”

Walter J, Steigemann W, Singh TP, Bartunik H, Bode W, and Huber R. 1982. “On the Disordered Activation Domain in Trypsinogen: Chemical Labelling and Low-Temperature Crystallography.” *Acta Crystallographica Section B Structural Crystallography and Crystal Chemistry* 38 (5). International Union of Crystallography (IUCr): 1462–72. doi:10.1107/s0567740882006153.

3.2 Bayesian Data Integration Questions Classic Study on Protease Self-Digest Kinetics

This section is based on the following publication:

[37] Tötsch, N & Hoffmann, D. (2020) Bayesian Data Integration Questions Classic Study on Protease Self-Digest Kinetics. *ACS Omega* **5**, 15162–15168
https://dx.doi.org/10.1021/acsomega.0c01109

ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

Bayesian Data Integration Questions Classic Study on Protease Self-Digest Kinetics

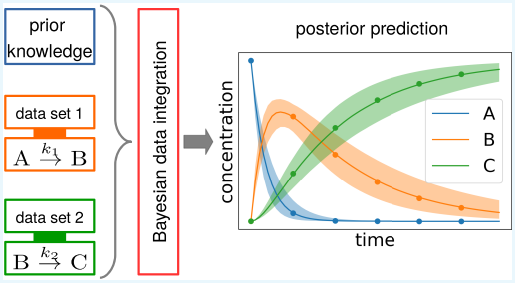
Niklas Tötsch* and Daniel Hoffmann

Cite This: *ACS Omega* 2020, 5, 15162−15168

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: We combine Bayesian data integration with kinetic modeling to rigorously identify reaction mechanisms. This approach forces models to be consistent not only with kinetic measurements but with all available information. We revisit a classic study on trypsin self-digest acceleration by colloidal silica. Bayesian data integration reveals that the mechanism suggested in that study is inconsistent with its presented data. We propose an improved hypothesis. However, the detailed mechanism of the surface reaction cannot be inferred from the available data.



INTRODUCTION

While the consistency of the narrow kinetic data with a hypothetic kinetic model is a necessary condition for the hypothesis to be correct, it is not sufficient proof. A more rigorous test would employ a broader set of heterogeneous data that can be used to reject plausible but false hypotheses. In this manuscript we employ such a more rigorous approach, based on the Bayesian integration of heterogeneous data sets, to test the reaction mechanisms. In this way, we show that not all plausible models are in agreement with the available data, which effectively reduces the number of possible mechanisms.

We aim to improve the mechanistic understanding of protease autolysis on surfaces. It is known that some important proteases undergo autolysis, i.e., cannibalistic self-digest, in solution.[1,2] This degradation can be accelerated. For instance, trypsin or α-chymotrypsin self-digest faster in the presence of glass or silica surfaces.[3] Similarly, savinase autolytic deactivation is accelerated by hydrophobic particles, showing that the phenomenon is not limited to trypsin-like proteases.[4]

Accelerated autolysis may often be an experimental nuisance. However, it could also be desirable in medical or technological applications as a means to inhibit proteases quickly and irreversibly. Thus, not surprisingly, there is currently renewed interest in controlling the activity by accelerating its self-digest. In 2015, Lv et al. successfully accelerated the autolytic trypsin deactivation by adding anionic polymers, such as dextran sulfate.[5] The irreversible inhibition of trypsin achieved by dextran sulfate in this way was 200 times more effective than that by a competitive trypsin inhibitor from soybean.

Gilles et al. observed substoichiometric inhibition of kallikrein and trypsin by several random copolymers in 2017.[6] Subsequent work by Smolin on trypsin and two polymers from Gilles' library confirmed the accelerated self-digest.[7] Such polymers, built from a library of monomers designed for protein surface recognition, might in the future allow to specifically accelerate protease self-digest or even protease−substrate digest. The more recent studies also provided the ultimate proof of autolysis, e.g., by showing a much faster build-up of autolysis products (P) in comparison to polymer-free control solutions.[5,7]

In the present work, we focus on the kinetic aspects of the accelerated self-digest; therefore, we revisit a detailed experimental study by Johnson and Whateley of the kinetics of acceleration of trypsin self-digest in the presence of silica surfaces.[8] We will refer to this publication as JW1981 in the remaining text. In agreement with their earlier work,[3] JW1981 observed that the relative deactivation is identical for three different starting concentrations of trypsin if silica is present (Figure 1), i.e., the reaction is of the first-order. They claimed that the bulk reaction (BR) of trypsin autolysis has a reaction order of 2.

JW1981 assumed that trypsin adsorbed on the silica surface (TS) is more susceptible to digestion by unbound trypsin (T) due to conformational changes. Thus, if the surface is fully covered, deactivation rates only depend on the trypsin concentration in the supernatant and therefore are first-order w.r.t. trypsin concentration. Although the model proposed by JW1981 is rather coarse-grained (the authors state that their theory "does not take into account many factors"), the

Received: March 12, 2020
Accepted: June 2, 2020
Published: June 19, 2020

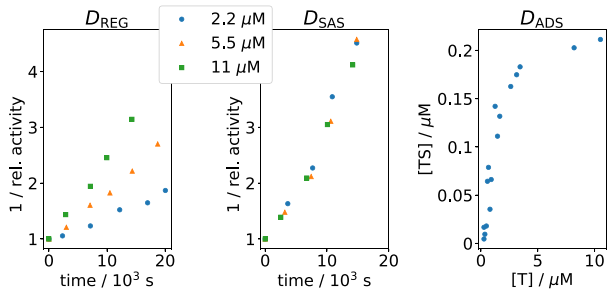


Figure 1. Trypsin deactivation and adsorption. Left: deactivation for different starting concentrations of trypsin $[T]_0$, 2 mM EDTA, no silica, i.e., regular self-digest (D_{REG}). Center: deactivation in the presence of 10 mg/L colloidal silica, i.e., surface-accelerated self-digest (D_{SAS}). Right: adsorption of trypsin on silica (D_{ADS}); 2 mM benzamidine, 2 mM $CaCl_2$, assuming 10 mg/L colloidal silica. We digitized and reformatted data from JW1981 Figures 4 and 8, respectively. Refer to Supporting Section S1 for details.

hypothesis (Scheme 1a) is *a priori* plausible. Since one of the reactants is adsorbed, whereas the other one is not, it resembles the Eley–Rideal (ER) mechanism from heterogeneous catalysis. Increased susceptibility of TS compared to that of trypsin (T) would be reflected in a rate constant of the ER mechanism (k_{ER}), which is larger than the rate constant of the bulk reaction (k_{BR}). More details, including the system of differential equations, can be found in the Supporting Information (Section S2.2). The adsorption of T to a binding site on the silica surface (S) is governed by the association rate (k_a) and dissociation rate (k_d). Their ratio determines the binding constant $K_a = \frac{k_a}{k_d}$.

JW1981's considerations regarding the underlying mechanism were based solely on the data of surface-accelerated self-digest (D_{SAS}), although they also present data on regular self-digest (D_{REG}) and an adsorption isotherm (D_{ADS}). JW1981 argued that at the relevant trypsin concentrations "the bulk second-order process is much slower than the surface reaction" because the overall reaction rate in D_{SAS} is of first-order, which is the same order they expected for the hypothesized surface reaction (SR). However, they did not test whether their proposed mechanism and their data were consistent with the assumption that BR is much slower than SR. Moreover, JW1981 assumed a constant coverage of the silica surfaces, but they did not crosscheck the trypsin concentrations in the kinetic experiments against their own D_{ADS} data.

JW1981 provided rich, heterogeneous data and a testable hypothesis on the basis of theoretical arguments and chemical intuitions. However, JW1981 did not fully exploit their own data to quantitatively test their hypothesis, possibly because of technical difficulties in solving the corresponding ordinary differential equation (ODE) system.

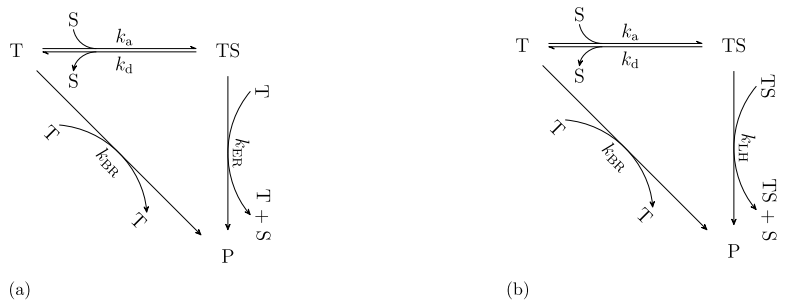
This setting is an excellent test case for a kinetic modeling approach that integrates all relevant data. Here, we demonstrate this with a Bayesian data integration (BDI) that implements a quantitative kinetic model of JW1981's hypothesis and tests it with the full, heterogeneous set of data from JW1981 (D_{REG}, D_{SAS}, D_{ADS}).

Another practical benefit of an explicit, quantitative treatment is the necessity to write down all chemical equations and to formulate underlying assumptions, some of which may not be obvious at first glance. We discuss nonobvious or implicit assumptions of our models in the Supporting Information (Section S5). Furthermore, an explicit, quantitative treatment can reveal flawed hypotheses. However, care must be taken not to introduce additional errors in the derivation of the ODE system. Fortunately, computational tools like `ChemPy` for constructing ODE systems from chemical equations, or `pyodesys` for solving ODE systems,[9] as well as our own additions presented in this work (BDI for ODE systems) streamline this process. The only input required for the kinetic models is the chemical equations.

We aim at the inference of kinetic parameters, given the input data from various sources, and a hypothetical kinetic model. The inferred parameters can then be used to predict the expected behavior of the modeled system. In particular, this allows us to compare the predicted and the observed behavior and thus to potentially reject hypothetical kinetic models. Moreover, multiple combinations of parameters may yield indistinguishable results in kinetic models, which match the experimental data. Broad parameter distributions indicate underdetermined models. It is therefore crucial to assess the uncertainty of the inferred parameters, given the uncertainties in the input data and their propagation through the inferential process. Thus, to account for uncertainties quantitatively, we express in BDI all quantities in terms of probability distributions.

Each data source, e.g., a thermodynamic or kinetic experiment, contributes an individual branch to the overall probabilistic model, with the sole constraint that each branch must share at least one parameter with another branch. This is a natural way of combining a heterogeneous set of such

Scheme 1. Chemical Equations for the Bulk Reaction (BR) of Tryptic Autolysis, as well as Suggested Surface Reaction (SR), with Trypsin (T), Binding Site on the Silica Surface (S), and Degradation Product (P)[a]



[a](a) JW1981's Eley–Rideal (ER) like mechanism and (b) alternative Langmuir–Hinshelwood (LH) like mechanism.

https://dx.doi.org/10.1021/acsomega.0c01109
ACS Omega 2020, 5, 15162−15168

branches, feeding the modeling process from diverse sources of data.

In general, a hypothetical kinetic mechanism will be tested more thoroughly if the inference of its parameters is based on a data set that is diverse. While there might be many mutually exclusive hypotheses explaining parts of the available data, there will be much fewer, ideally only one hypothesis, compatible with all observations.

The principal Bayesian approach to infer mechanistic model parameters from a single data set has been established several years ago.[10] More recently, Choi and Tomczak independently inferred enzyme kinetic constants with Bayesian tools.[11,12] BDI for heterogeneous data sources has been applied successfully to metabolic networks,[13] toxicokinetics,[14] and system biology.[15]

To confirm that our BDI method can indeed infer parameters for chemical kinetic models and discard or support models based on their consistency with data, we present an example with synthetic data in the Supporting Information (Section S7). This proof of concept also serves as a reference for readers interested in the implementation and as a tutorial for potential users. It can be found at https://github.com/niklastoe/kineticmodel_bdi. In the following sections, we focus on the modeling of trypsin autolysis in the presence of silica surfaces, and we demonstrate that some of the published claims about the process are no longer tenable if all data are considered.

RESULTS AND DISCUSSION

JW1981's Hypothesis is Compatible with D_{SAS} Data Alone. At first, all required parameters, those shown in Scheme 1a as well as the initial concentration of the binding sites on the silica surface $[S]_0$, are inferred from D_{SAS} alone. The experimental data are reproduced seemingly well (Figure 2a), which is necessary but not sufficient for the validity of the hypothesis assumed by JW1981. Remarkably, marginal

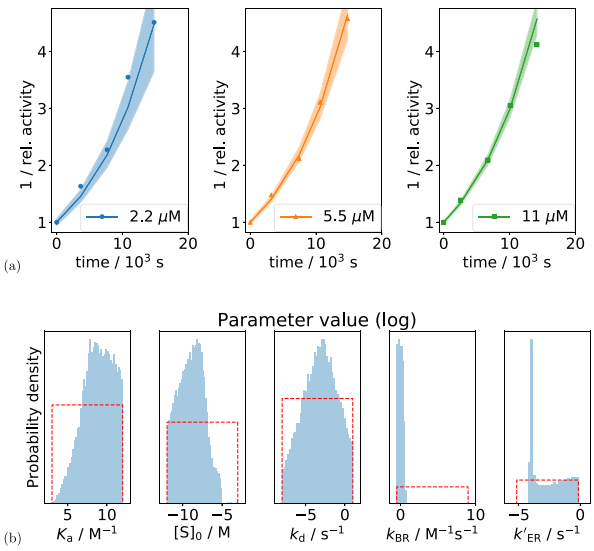


Figure 2. Results of kinetic modeling and Bayesian inference for JW1981's hypothesis. (a) D_{SAS} (markers as in Figure 1) for different $[T]_0$ are mostly within 16−84% uncertainty intervals of the ER model (contour). (b) Priors (red, dashed) and marginalized posterior distributions inferred from D_{SAS} (blue). Refer to Supporting Section S6.2.2 for details.

posteriors for most parameters indicate large uncertainties of parameter values (Figure 2b). Due to the limited amount of kinetic data and lack of constraints for the parameters, this model is underdetermined, i.e., there are many combinations of parameters for the modeled ER mechanism that are consistent with the same kinetic data.

Alternative Langmuir−Hinshelwood-Like and JW1981 Mechanisms Explain Observations Equally Well. Underdetermination does not only allow many parameter value combinations to explain the kinetic data but even permits different models. For instance, Figure 3 shows

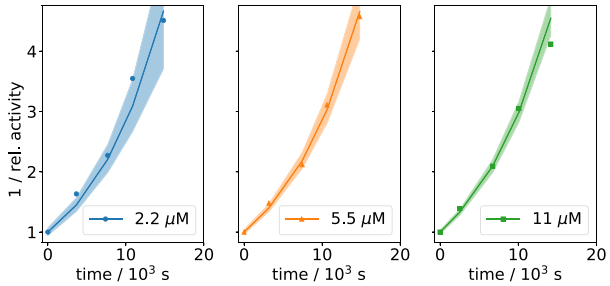


Figure 3. D_{SAS} (markers as in Figure 1) for different $[T]_0$ are within 16%−84% uncertainty intervals of the LH model (contour).

that D_{SAS} is also compatible with a Langmuir−Hinshelwood (LH) like mechanism (Scheme 1b). LH is another textbook mechanism from heterogeneous catalysis. In contrast to ER, it assumes that both molecules are adsorbed before the reaction (see Supporting Section S2.3 for more details and the system of differential equations). A possible interpretation of an accelerated self-digest, i.e., increased rate constant $k_{LH} > k_{BR}$, could be local enrichment on the surface. Considering only D_{SAS}, both ER and LH have similar explanatory powers, prohibiting the rejection of either one mechanism. Like JW1981's proposed mechanism, LH is a coarse-grained description lacking, e.g., accurate treatment of local concentrations or surface diffusion.

Using More Data from JW1981 Challenges JW1981's Hypothesis. For inference of all reaction parameters of the two models considered so far, ER and LH, we need not only D_{SAS} but also other experimental data. Namely, we need D_{REG} for the inference of the regular self-digest rate in solution, k_{BR} (Supporting Section S4), and D_{ADS} for the description of the concentration of silica surface binding sites $[S]_0$ and the association constant of trypsin and the binding sites, K_a (Supporting Section S2.1). These three parameters, k_{BR}, $[S]_0$, and K_a, are required for a quantitative kinetic model of the total reaction with simultaneous bulk and surface reactions, BR and SR (Scheme 1a,b), respectively.

D_{REG} and D_{SAS} Imply Different k_{BR}. We estimated the total reaction rates in the presence or absence of silica surfaces (Figure S4). Silica can increase reaction rates by more than 100% for $[T]_0$ = 2.2 μM. For $[T]_0$ = 5.5 μM and especially $[T]_0$ = 11.0 μM, the presence of silica surfaces does not lead to much faster overall reaction rates. Thus, for these concentrations, one cannot neglect BR because its speed could be of the same order of magnitude as that of SR.

JW1981 described regular autolysis as a second-order reaction. In our first model, we describe the regular bulk autolysis BR with a single-step bimolecular reaction (Supporting Section S4). It seems plausible that the presence of silica

https://dx.doi.org/10.1021/acsomega.0c01109
ACS Omega 2020, 5, 15162−15168

does not influence the bulk reaction rate k_{BR}. Joint modeling of $D = \{D_{REG}, D_{SAS}\}$ is equivalent to inferring k_{BR} from D_{REG} alone and then using it as a prior for modeling D_{SAS} (Supporting Section S6) as mentioned previously in the paper.

If we do this, the agreement between model predictions and the observed D_{SAS} deteriorates (Figure 4, left). Uncertainty

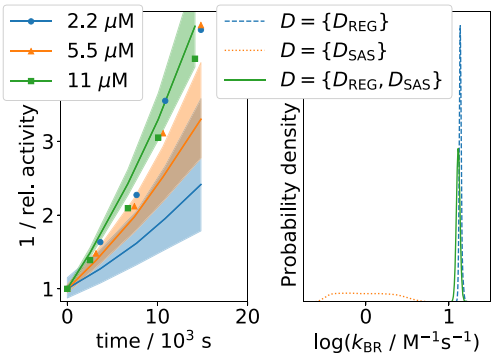


Figure 4. Modeling with joint data of bulk and surface reactions, D_{REG} and D_{SAS}. Left: comparison of experimental D_{SAS} data (markers) and 16−84% uncertainty intervals (contour) of the ER model based on $D = \{D_{REG}, D_{SAS}\}$. Right: marginalized posterior distribution for k_{BR} inferred for the ER model from $D = \{D_{SAS}\}$ or $D = \{D_{REG}\}$ individually, or inferred from joint $D = \{D_{REG}, D_{SAS}\}$.

intervals are broader than if using D_{SAS} data alone (compare to Figure 2a). The model predictions systematically deviate from the experimental data for all $[T]_0$. Specifically, the model based on the joint D_{REG} and D_{SAS} data predicts that higher concentrations deactivate faster, i.e., the model is not first-order. Obviously, this is incompatible with the experimental data. If k_{BR} is required to match D_{REG}, BR contributes substantially to the overall degradation in this model if silica is present, leading to a reaction order greater than 1.

If we model D_{REG} and D_{SAS} separately, the marginalized posterior distributions for k_{BR} do not overlap (Figure 4, right). This means that JW1981's model cannot explain both data sets with a coherent interval of model parameter k_{BR}. Barring measurement errors in JW1981's data, we are confident that a single model should be able to explain both D_{REG} and D_{SAS} because their experimental setup is identical. Consequently, we reject both the ER- and the LH-like mechanisms (similar results to Figure 4, not shown) as defined in Scheme 1a,1b.

Contrary to the Hypothesis, Surface Coverage is Not Constant. JW1981 claimed that surfaces are fully covered in their kinetic experiments. Surface coverage (ϕ) would be approximately constant, leading to a first-order reaction assuming an ER-like mechanism (Supporting Section S2.2). ϕ is assumed to follow the Langmuir isotherm.

$$\phi = \frac{K_a[T]}{1 + K_a[T]} = \frac{[TS]}{[S]_0} \quad (1)$$

Rearranging eq 1 one obtains a function of ϕ dependent on the starting concentration $[T]_0$. We calculated the expected ϕ for the starting concentrations of trypsin in JW1981's kinetic experiments, with our BDI approach propagating the uncertainty of the parameters K_a and $[S]_0$ inferred from D_{ADS} (Figure 5). Contrary to JW1981's statement, we find that ϕ is not approaching full surface coverage. Even the largest starting concentration of trypsin of 11 μM has a 98% probability for $\phi < 0.9$. More importantly, ϕ differs

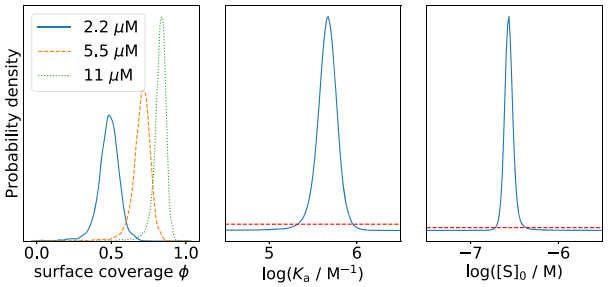


Figure 5. Posterior prediction of equilibrium ϕ for different $[T]_0$ (left). Marginal posterior distributions of K_a (middle) and $[S]_0$ (right) have been inferred from D_{ADS}. Priors are shown as red dashed lines.

substantially for the three trypsin concentrations. This finding contradicts the above hypothesis that the amount of substrate, i.e., adsorbed trypsin [TS], is constant, which was a crucial part of JW1981's explanation of a first-order reaction.

We could think of two possible reservations against using parameters inferred from D_{ADS} to estimate the concentrations in kinetic experiments.

First, the differences in experimental conditions (presence of Ca^{2+} and benzamidine in the adsorption experiments but not the kinetic measurements) may distort K_a and $[S]_0$. JW1981 rejected the notion that benzamidine could influence the adsorption but do not comment on Ca^{2+}. If Ca^{2+} decreases K_a (see Supporting Section S2.1), JW1981's assumption of a constant ϕ might be satisfied in the kinetic experiments.

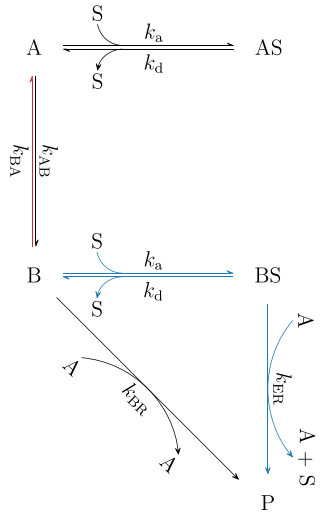
Second, the Langmuir isotherm cannot reproduce D_{ADS} accurately as we will demonstrate later. This could be caused by a violation of one or more of the assumptions made by the Langmuir isotherm or due to a systematic error in D_{ADS}. JW1981 chose to use the Langmuir isotherm, and we are not aware of another isotherm leading to a better description.

Either case could lead to wrong posterior distributions of K_a and $[S]_0$, resulting in wrong estimates of the surface coverage; however, these are merely hypothetical scenarios. We find that K_a would need to be at least an order of magnitude larger to ensure constant surface coverage (Supporting Section S3). Thus, JW1981's claim that surface coverage is constant for the given concentrations of trypsin is unsupported at best. Note that even if applying parameters inferred from D_{ADS} to concentrations in kinetic experiments was inadmissible and surface coverage was indeed constant in kinetic experiments, BDI of $D = \{D_{REG}, D_{SAS}\}$ is sufficient to reject JW1981's overall hypothesis.

Including Transition between Two Trypsin Conformations A and B Allows to Explain All Data Holistically. Kunitz and Northrop proposed in 1934 that trypsin autolysis involves two conformations of the protease.[1] Conformation B is susceptible to autolysis by A, but not the other way around. If this conversion is slow/rate-determining, the reaction order of regular autolysis can be below 2.[2,16] JW1981 mentioned the influence of Ca^{2+} on trypsin's conformation and its impact on the susceptibility but do not consider two forms A and B in their hypothesis for surface-accelerated self-digest. A closer inspection of D_{REG} shows that its reaction order is most likely below 2 (Supporting Section S4.3). To test this possibility further, we therefore adapt our kinetic model by treating trypsin as two forms A and B (Scheme 2, the system of differential equations can be found in Supporting Section S2.4).

https://dx.doi.org/10.1021/acsomega.0c01109
ACS Omega 2020, 5, 15162−15168

Scheme 2. Chemical Equations for the Interconversion of Two Trypsin Conformations A and B and the Corresponding BR- or ER-Like Mechanism of Degradation



Except for the autolytic step, we assume A and B to be equivalent (see also Supporting Section S5.6): A and B adsorb equally to silica, i.e., adsorption/desorption rate constants are identical for A and B, which keeps the model simple and avoids inflating the parameter space.

This new, refined model reproduces the heterogeneous data set $D = \{D_{REG}, D_{SAS}, D_{ADS}\}$ accurately (Figure 6a). The total

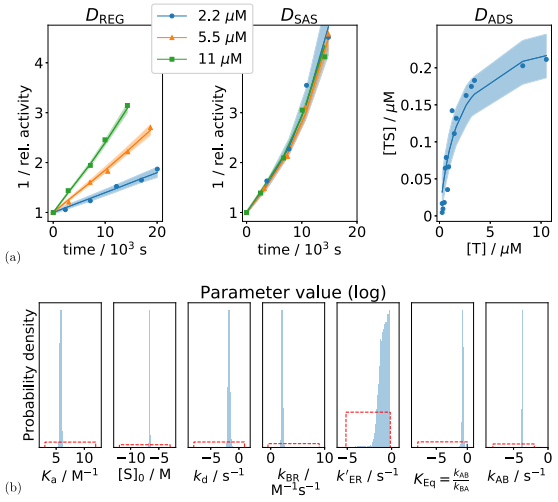


Figure 6. Results of kinetic modeling and Bayesian inference for the model presented in Scheme 2. (a) Posterior predictive check for data sets from all experiments. (b) Priors (red, dashed) and marginalized posterior distributions inferred from $D = \{D_{REG}, D_{SAS}, D_{ADS}\}$ (blue). Refer to Supporting Section S6.2.2 for details.

reaction order in D_{SAS} is appropriate, although the surface coverage is not constant for the examined trypsin concentrations. SR does not govern the reaction rate. It simply accelerates the degradation of B (shown in blue in Scheme 2). Consequently, conversion from A to B becomes the rate-determining step, causing the pseudo-first-order reaction rate. A and B are no longer in a steady state because the conversion B → A (shown in red in Scheme 2) is slower than the degradation of B.

The posterior distributions of the parameters are much more narrow if all three available data sets are used for the inference (Figure 6b). Only k_{ER} is broadly distributed because it is not rate-determining. As long as k_{ER} is much higher than the conversion rate from A to B (k_{AB}), the modeled digestion rates do not change. Independent measurements of k_{AB} and the equilibrium constant of the two trypsin conformations $K_{Eq} = \frac{k_{AB}}{k_{BA}}$ would test our hypothesis.

JW1981 compared the effect of silica to the one by Ca^{2+} which increases the resistance to autolysis. Both lower the reaction order but silica increases the overall reaction rate, whereas Ca^{2+} decreases it. A simple explanation for this phenomenon could be that silica accelerates the degradation of B, whereas Ca^{2+} slows the conversion A → B. In both cases, the conversion becomes the rate-determining step.

We emphasize that we do not claim that the last model is necessarily correct. We only state that it is one model that is consistent with the available data and with the literature (ingrained explicitly in the priors or not). The only conclusion we draw is that silica accelerates the degradation of B to a point where it is no longer rate-determining. The exact mechanism of acceleration cannot be determined based on the available data. For example, we cannot determine if increased susceptibility of adsorbed trypsin or an increase in local concentration makes SR faster than BR. We refined the LH mechanism discussed earlier with conformations A and B. As expected, this mechanism explains the full heterogeneous data set just as well as by the mechanism in Scheme 2 (results not shown). Studying SR in detail would probably be easiest if it was the rate-determining step.

CONCLUSIONS

We could show the inadequacy of the mechanism proposed by JW1981, as it is incompatible with JW1981's full data on the surface-accelerated trypsin self-digest, regular self-digest, and adsorption, all integrated with BDI. An extended model, describing trypsin in two forms, form A and a more quickly degraded form B, leads to a pseudo-first-order reaction that explains all the available data. This is true irrespective of whether we assume in the model that silica accelerates the process in an ER- or LH-like mechanism. We therefore conclude that the data presented in JW1981 is insufficient to determine how silica accelerates trypsin self-digest. In addition, we stress that due to the limitations of the data available from JW1981, especially the complete lack of error estimates, the reliability of any model is also limited.

Recently, Zumbro et al. found that polymers increase the local concentration of unbound proteins in a protein-rich phase if these proteins attract each other.[17] This increase might lead to a faster self-digest without adsorbing the proteases involved in autolysis, i.e., it would be quite different from the LH-like mechanism proposed in Scheme 1b. Future studies trying to identify details of SR might need to consider this third variant. Describing this phenomenon with a kinetic model is not straightforward since adsorption and aggregation close to the surface are themselves dependent on the surface coverage and local concentration close to the surface.

The case of JW1981 has also demonstrated the importance of data integration in modeling. Modeling D_{SAS} individually led to severely underdetermined models that were rejected after

https://dx.doi.org/10.1021/acsomega.0c01109
ACS Omega 2020, 5, 15162−15168

the application of BDI to the complete data $D = \{D_{REG}, D_{SAS}, D_{ADS}\}$. The general approach proposed here, i.e., BDI for heterogeneous data sets, will be useful for other applications in chemical kinetics.

METHODS

Kinetic Modeling Setup. We used `ChemPy` to create systems of ODEs from chemical equations[9] and `pyodesys` to integrate the systems of ODEs numerically with `CVODE`.[18,19] Although numerical integration comes at the cost of slower posterior evaluations during inference, we believe that in many scenarios deploying models rapidly, as facilitated by the described set of tools, is more important than reduced computational cost, especially if there are numerous hypotheses to test.

Experimental Uncertainties. JW1981 does not contain error bars. Our probabilistic kinetic models included experimental uncertainties σ_{kin} (for D_{REG} and D_{SAS}) and σ_{ads} (for D_{ADS}) as nuisance parameters (Supporting Section S6.1).

Sampling. For numerical Bayesian inference, we used the ensemble sampler for Markov chain Monte Carlo proposed by Goodman and Weare as implemented in `emcee`.[20,21] Sampling was parallelized with `pathos`.[22,23]

For the representative sampling of the parameter space, we used 500 MCMC walkers. We stopped sampling when autocorrelation times τ converged (less than 2% change within 2500 steps) for all parameters and the run time exceeded the largest τ at least 50 times. After discarding the first $5{\cdot}\tau$ steps as burn-in, we evaluated only every τ-th step, i.e., at least 22 500 independent samples.

Convergence was checked with `PyMC`'s implementation of the Gelman–Rubin diagnostics ($R_c < 1.01$).[24−26]

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.0c01109.

> Details on how experimental data were extracted and reformatted from JW1981 (Section S1); theoretical background for surface reactions (Section S2); requirements for constant surface coverage (Section S3); regular autolysis and inference of its reaction order in the experimental data (Section S4); implicit assumptions made in our model (Section S5); and theoretical background of Bayesian inference, Bayesian data integration, and reasoning behind chosen priors (Section S6) (PDF)

AUTHOR INFORMATION

Corresponding Author

Niklas Tötsch − *Bioinformatics and Computational Biophysics, Universität Duisburg-Essen, 45141 Essen, Germany;* orcid.org/0000-0001-7656-1160; Email: niklas.toetsch@uni-due.de

Author

Daniel Hoffmann − *Bioinformatics and Computational Biophysics, Universität Duisburg-Essen, 45141 Essen, Germany;* orcid.org/0000-0003-2973-7869

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.0c01109

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Funding from Deutsche Forschungsgemeinschaft through Project CRC1093/A7 is gratefully acknowledged.

REFERENCES

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

(2) 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.

(3) Johnson, P.; Whateley, T. The effect of glass and silica surfaces on trypsin and α-chymotrypsin kinetics. *Biochim. Biophys. Acta, Enzymol.* **1972**, *276*, 323−327.

(4) Maste, M. C.; Rinia, H. A.; Brands, C. M.; Egmond, M. R.; Norde, W. Inactivation of a subtilisin in colloidal systems. *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.* **1995**, *1252*, 261−268.

(5) Lv, Y.; Zhang, J.; Wu, H.; Zhao, S.; Song, Y.; Wang, S.; Wang, B.; Lv, G.; Ma, X. A protease inhibition strategy based on acceleration of autolysis. *Chem. Commun.* **2015**, *51*, 5959−5962.

(6) Gilles, P.; Wenck, K.; Stratmann, I.; Kirsch, M.; Smolin, D. A.; Schaller, T.; De Groot, H.; Kraft, A.; Schrader, T. High-Affinity Copolymers Inhibit Digestive Enzymes by Surface Recognition. *Biomacromolecules* **2017**, *18*, 1772−1784.

(7) Smolin, D. A. Maßgeschneiderte Affinitätspolymere als Katalysatoren für die Autoproteolyse von Trypsin. Ph.D. Thesis, University of Duisburg-Essen, 2018.

(8) Johnson, P.; Whateley, T. L. The effects of surface and macromolecular interactions on the kinetics of inactivation of trypsin and alpha-chymotrypsin. *Biochem. J.* **1981**, *193*, 285−294.

(9) Dahlgren, B. ChemPy: A package useful for chemistry written in Python. *J. Open Source Softw.* **2018**, *3*, No. 565.

(10) Vyshemirsky, V.; Girolami, M. A. Bayesian ranking of biochemical system models. *Bioinformatics* **2008**, *24*, 833−839.

(11) Choi, B.; Rempala, G. A.; Kim, J. K. Beyond the Michaelis-Menten equation: Accurate and efficient estimation of enzyme kinetic parameters. *Sci. Rep.* **2017**, *7*, No. 17018.

(12) Tomczak, J. M.; Węglarz-Tomczak, E. Estimating kinetic constants in the Michaelis-Menten model from one enzymatic assay using Approximate Bayesian Computation. *FEBS Lett.* **2019**, *593*, 2742−2750.

(13) Liebermeister, W.; Klipp, E. Bringing metabolic networks to life: integration of kinetic, metabolic, and proteomic data. *Theor. Biol. Med. Model.* **2006**, *3*, No. 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.

(16) Yon, J. Action du chlorure de sodium sur l'autolyse de la trypsine. *Biochim. Biophys. Acta* **1959**, *31*, 75−85.

(17) Zumbro, E.; Witten, J.; Alexander-Katz, A. Computational Insights into Avidity of Polymeric Multivalent Binders. *Biophys. J.* **2019**, *117*, 892−902.

(18) Dahlgren, B. pyodesys: Straightforward numerical integration of ODE systems from Python. *J. Open Source Softw.* **2018**, *3*, No. 490.

(19) Hindmarsh, A. C.; Brown, P. N.; Grant, K. E.; Lee, S. L.; Serban, R.; Shumaker, D. E.; Woodward, C. S. SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers. *ACM Trans. Math. Softw.* **2005**, *31*, 363−396.

(20) Goodman, J.; Weare, J. Ensemble samplers with affine invariance. *Commun. Appl. Math. Comput. Sci.* **2010**, *5*, 65−80.

(21) Foreman-Mackey, D.; Hogg, D. W.; Lang, D.; Goodman, J. emcee: The MCMC Hammer. *Publ. Astron. Soc. Pac.* **2013**, *125*, 306−312.
(22) McKerns, M. M.; Aivazis, M. A. G. pathos: A Framework for Heterogeneous Computing. http://trac.mystic.cacr.caltech.edu/project/pathos.
(23) McKerns, M. M.; Strand, L.; Sullivan, T.; Fang, A.; Aivazis, M. A. G. Building a Framework for Predictive Science. 2012, arXiv:1202.1056. arXiv.org e-Print archive. https://arxiv.org/abs/1202.1056.
(24) Gelman, A.; Rubin, D. B. Inference from Iterative Simulation Using Multiple Sequences. *Stat. Sci.* **1992**, *7*, 457−472.
(25) Brooks, S. P.; Gelman, A. General Methods for Monitoring Convergence of Iterative Simulations. *J. Comput. Graph. Stat.* **1998**, *7*, 434−455.
(26) Salvatier, J.; Wiecki, T. V.; Fonnesbeck, C. Probabilistic programming in Python using PyMC3. *PeerJ Comput. Sci.* **2016**, *2*, No. e55.

https://dx.doi.org/10.1021/acsomega.0c01109
ACS Omega 2020, 5, 15162−15168

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 $[\mathrm{T}]_t + [\mathrm{TS}]_t$.

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

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.

$$\mathrm{T} + \mathrm{S} \underset{k_\mathrm{d}}{\overset{k_\mathrm{a}}{\rightleftharpoons}} \mathrm{TS} \tag{S1}$$

where:

S = unoccupied binding site on the silica surface

TS = trypsin adsorbed on the silica surface

k_a = association rate

k_d = dissociation rate

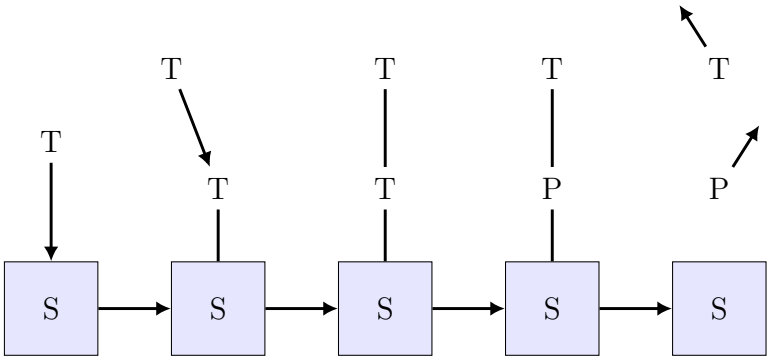
Koutsopoulos et al. estimated that a tightly packed monolayer of trypsins would lead to a coverage of 2.1 $\mathrm{mg\,m^{-2}}$.[2] On silica slides, they observed a maximum coverage of 1.1 $\mathrm{mg\,m^{-2}}$ whereas Berg et al. observed a peak around 2 $\mathrm{mg\,m^{-2}}$.[3] Syton X30, the colloidal silica used by JW1981, has a specific surface of 242 $\mathrm{m^2\,g^{-1}}$.[1] Multiplied by Koutsopoulos' estimate of the density of a tightly packed monolayer, Syton X30 could bind up to 0.51 $\frac{\mathrm{mg\ T}}{\mathrm{mg\ SiO_2}}$. 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 $K_\mathrm{a}{=}\frac{k_\mathrm{a}}{k_\mathrm{d}}$. Berg's experiments approach full monolayer coverage, i.e. $\phi \approx 1$, with an initial trypsin concentration $[\mathrm{T}]_0 = 0.43$ µM. 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 CaCl2, 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).



Scheme S1: Eley-Rideal (ER)-like mechanism

$$\mathrm{TS} + \mathrm{T} \overset{k_{\mathrm{ER}}}{\rightarrow} \mathrm{T} + \mathrm{S} + \mathrm{P} \tag{S2}$$

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

$$r_{\mathrm{ER}}^{\mathrm{surface}} = \left(\frac{d[\mathrm{P}]^{\mathrm{surface}}}{dt}\right)_{\mathrm{ER}} = C_S k''_{ER} \phi[\mathrm{T}] \tag{S3}$$

where:

C_S = binding site density on the surface, $\mathrm{mol\,m^{-2}}$

k''_{ER} = rate constant, $\mathrm{M^{-1}\,s^{-1}}$

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 $r_{\mathrm{ER}}^{\mathrm{solution}}$ (in $\mathrm{M^{-1}\,s^{-1}}$) is defined by

$$r_{\mathrm{ER}}^{\mathrm{solution}} = \left(\frac{d[\mathrm{P}]^{\mathrm{solution}}}{dt}\right)_{\mathrm{ER}} = A\, r_{\mathrm{ER}}^{\mathrm{surface}} = AC_S k''_{ER}\phi[\mathrm{T}] = k'_{ER}\phi[\mathrm{T}] = k'_{ER}\frac{[\mathrm{TS}]}{[\mathrm{S}]_0}[\mathrm{T}] = k_{\mathrm{ER}}[\mathrm{TS}][\mathrm{T}] \tag{S4}$$

where:

A = silica surface per volume, $\mathrm{m}^2\,\mathrm{L}^{-1}$

$k'_{ER} = AC_S k''_{ER}$

$k_{\mathrm{ER}} = \frac{k'_{ER}}{[\mathrm{S}]_0}$

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

$$\frac{d[\mathrm{P}]}{dt} = k_{\mathrm{BR}}[\mathrm{T}]^2 + k_{ER}[\mathrm{T}][\mathrm{TS}] \tag{S5}$$

$$\frac{d[\mathrm{S}]}{dt} = -k_{\mathrm{a}}[\mathrm{S}][\mathrm{T}] + k_{\mathrm{d}}[\mathrm{TS}] + k_{\mathrm{ER}}[\mathrm{T}][\mathrm{TS}] \tag{S6}$$

$$\frac{d[\mathrm{T}]}{dt} = -k_{\mathrm{BR}}[\mathrm{T}]^2 - k_{\mathrm{a}}[\mathrm{S}][\mathrm{T}] + k_{\mathrm{d}}[\mathrm{TS}] \tag{S7}$$

$$\frac{d[\mathrm{TS}]}{dt} = k_{\mathrm{a}}[\mathrm{S}][\mathrm{T}] - k_{\mathrm{d}}[\mathrm{TS}] - k_{\mathrm{ER}}[\mathrm{T}][\mathrm{TS}]. \tag{S8}$$

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

$$\omega_{[\mathrm{T}]}^{\mathrm{ER}} = \frac{d\ln r_{\mathrm{ER}}^{\mathrm{solution}}}{d\ln[\mathrm{T}]} = 2 - \phi. \tag{S9}$$

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

If ϕ 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.

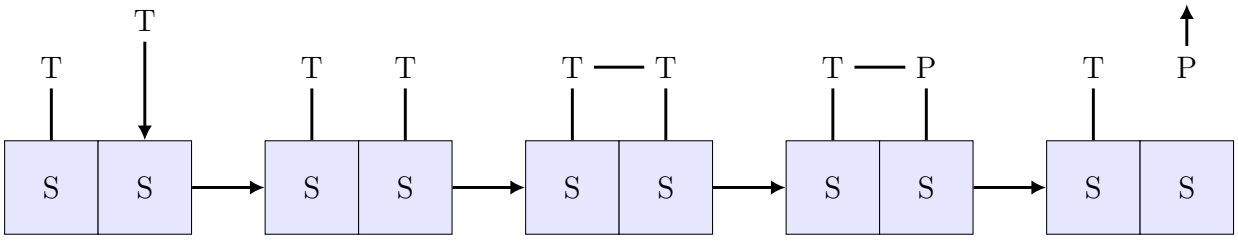
Eq. S2 assumes that the substrate is adsorbed, whereas the digester attacks from the bulk. In principle, these roles could also be reversed with an immobilized digester on the surface proteolysing substrates from the solution:

$$\mathrm{TS} + \mathrm{T} \overset{k_{\mathrm{ER}}}{\rightarrow} \mathrm{TS} + \mathrm{P} \tag{S10}$$

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).



Scheme S2: Langmuir-Hinshelwood (LH)-like mechanism

$$2\ \mathrm{TS} \overset{k_{\mathrm{LH}}}{\rightarrow} \mathrm{TS} + \mathrm{P} + S \tag{S11}$$

$$r_{\mathrm{LH}}^{\mathrm{surface}} = \left(\frac{d[\mathrm{P}]^{surface}}{dt}\right)_{LH} = C_S^2 k''_{LH}\phi\phi \tag{S12}$$

where:

k''_{LH} = rate constant, $\mathrm{m}^2\,\mathrm{M}^{-1}\,\mathrm{s}^{-1}$

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

where:

k'_{LH} = $AC_S^2 k''_{LH}$

k_{LH} = $\frac{k'_{LH}}{([\mathrm{S}]_0)^2}$

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

$$\frac{d[\mathrm{P}]}{dt} = k_{\mathrm{BR}}[\mathrm{T}]^2 + k_{\mathrm{LH}}[\mathrm{TS}]^2 \tag{S14}$$

$$\frac{d[\mathrm{S}]}{dt} = -k_{\mathrm{a}}[\mathrm{S}][\mathrm{T}] + k_{\mathrm{d}}[\mathrm{TS}] + k_{\mathrm{LH}}[\mathrm{TS}]^2 \tag{S15}$$

$$\frac{d[\mathrm{T}]}{dt} = -k_{\mathrm{BR}}[\mathrm{T}]^2 - k_{\mathrm{a}}[\mathrm{S}][\mathrm{T}] + k_{\mathrm{d}}[\mathrm{TS}] \tag{S16}$$

$$\frac{d[\mathrm{TS}]}{dt} = k_{\mathrm{a}}[\mathrm{S}][\mathrm{T}] - k_{\mathrm{d}}[\mathrm{TS}] - k_{\mathrm{LH}}[\mathrm{TS}]^2. \tag{S17}$$

If unbound and adsorbed trypsin are in equilibrium, the reaction order (ω) equals

$$\omega_{[\mathrm{T}]}^{\mathrm{LH}} = \frac{d \ln r_{\mathrm{LH}}^{\mathrm{solution}}}{d \ln[\mathrm{T}]} = 2 \cdot (1 - \phi). \tag{S18}$$

If ϕ approaches maximum coverage, product formation converges to a maximum since the maximum amount of available [TS] is capped by the number of binding sites $[\mathrm{S}]_0$. 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

$$\frac{d[\mathrm{A}]}{dt} = -k_{\mathrm{AB}}[\mathrm{A}] + k_{\mathrm{BA}}[\mathrm{B}] - k_{\mathrm{a}}[\mathrm{A}][\mathrm{S}] + k_{\mathrm{d}}[\mathrm{AS}] \tag{S19}$$

$$\frac{d[\mathrm{AS}]}{dt} = k_{\mathrm{a}}[\mathrm{A}][\mathrm{S}] - k_{\mathrm{d}}[\mathrm{AS}] \tag{S20}$$

$$\frac{d[\mathrm{B}]}{dt} = k_{\mathrm{AB}}[\mathrm{A}] - k_{\mathrm{BA}}[\mathrm{B}] - k_{\mathrm{BR}}[\mathrm{A}][\mathrm{B}] - k_{\mathrm{a}}[\mathrm{B}][\mathrm{S}] + k_{\mathrm{d}}[\mathrm{BS}] \tag{S21}$$

$$\frac{d[\mathrm{BS}]}{dt} = k_{\mathrm{a}}[\mathrm{B}][\mathrm{S}] - k_{\mathrm{d}}[\mathrm{BS}] - k_{\mathrm{ER}}[\mathrm{A}][\mathrm{BS}] \tag{S22}$$

$$\frac{d[P]}{dt} = k_{\mathrm{BR}}[\mathrm{A}][\mathrm{B}] + k_{\mathrm{ER}}[\mathrm{A}][\mathrm{BS}] \tag{S23}$$

$$\frac{d[\mathrm{S}]}{dt} = -k_{\mathrm{a}}[\mathrm{S}]([\mathrm{A}] + [\mathrm{B}]) + k_{\mathrm{d}}([\mathrm{AS}] + [\mathrm{BS}]) + k_{\mathrm{ER}}[\mathrm{A}][\mathrm{BS}] \tag{S24}$$

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

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

Setting $[\mathrm{T}]_0^{\mathrm{a}} = 11.0\ \mu M$ and $[\mathrm{T}]_0^{\mathrm{b}} = 2.2\ \mu M$, 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).

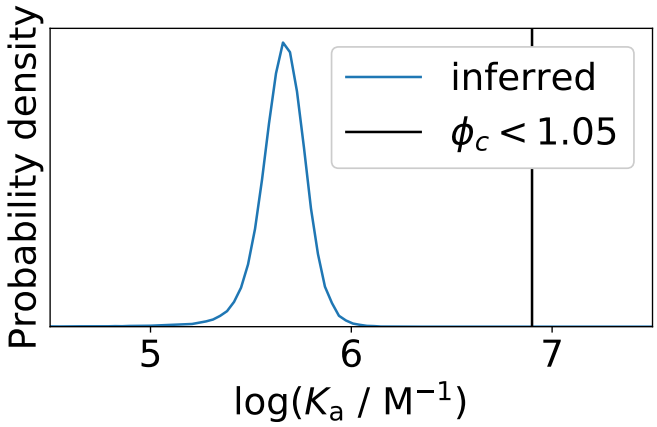


Figure S1: K_{a} would need to be substantially larger than implied by D_{ADS} for constant ϕ.

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.

$$2\ \mathrm{T} \overset{k_{\mathrm{BR}}}{\rightarrow} \mathrm{T} + \mathrm{P} \tag{S26}$$

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[4] but neglecting the digester-substrate complex is given by

$$\mathrm{A} \underset{k_{\mathrm{BA}}}{\overset{k_{\mathrm{AB}}}{\rightleftharpoons}} \mathrm{B} \tag{S27}$$

$$\mathrm{A} + \mathrm{B} \overset{k_{\mathrm{BR}}}{\rightarrow} \mathrm{A} + \mathrm{P} \tag{S28}$$

S4.3 Reaction order of regular autolysis is below 2

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

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

First, we calculate

$$r = \frac{[\mathrm{T}]_{t_2} - [\mathrm{T}]_{t_1}}{t_2 - t_1} \tag{S30}$$

from data on regular self-digest (D_{REG}) for the smallest timestep $t_2 - t_1$ available. We then perform Bayesian inference of ω and C (which is a constant of no further interest to us) with this probabilistic model:

$$
\begin{aligned}
\log r^{\text{exp}} &\sim N(\log r^{\text{model}}, \sigma) \\
\log r^{\text{model}} &\sim \omega \cdot \log([\text{T}]) + C \\
\omega &\sim \text{unif}(0.0, 3.0) \\
C &\sim \text{unif}(-20.0, 20.0) \\
\sigma &\sim \text{unif}(0.001, 10.0)
\end{aligned}
\tag{S31}
$$

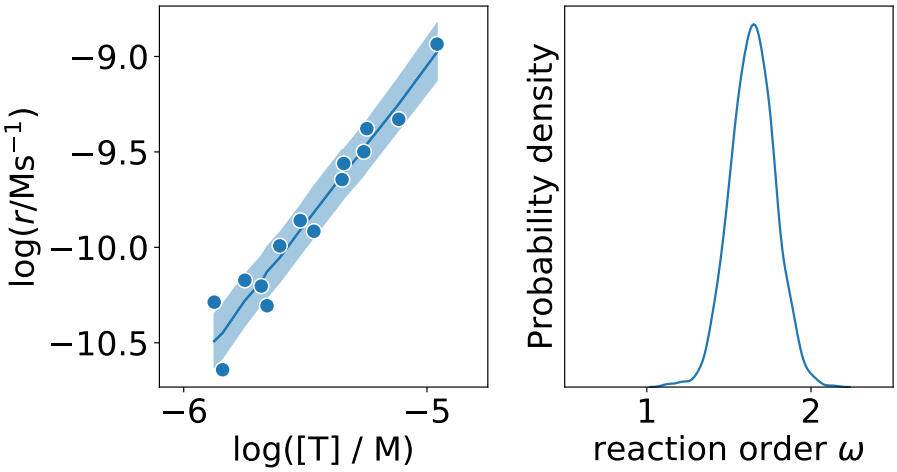


Figure S2: Logarithm of reaction rates r plotted against logarithms of trypsin concentrations in D_{REG} and posterior predictive check (left). Marginalized posterior distribution for ω (right).

S4.3.1 Inclusion of two trypsin states can explain D_{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_{REG} (Figure S3).

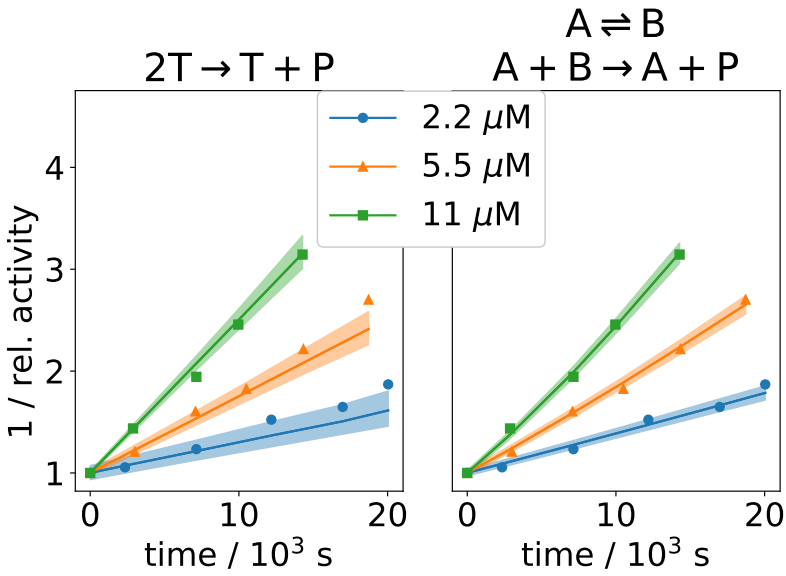


Figure S3: Posterior predictive check for two models of regular autolysis.

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

Reaction rates with and without silica differ strongly for $[\mathrm{T}]_0 = 2.2$ µM but not for $[\mathrm{T}]_0 = 11.0$ µM (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.[3] At first,

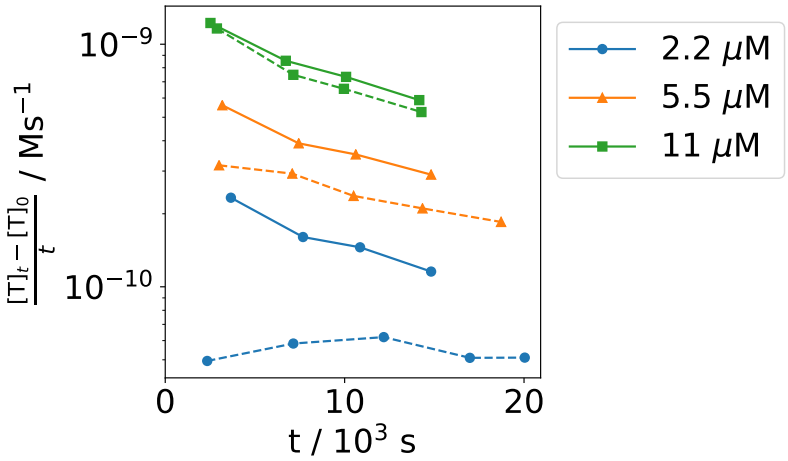


Figure S4: Experimental reaction rates for different trypsin starting concentrations if silica is absent (D_{REG}, dashed) or present (D_{SAS}, solid).

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}.[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.[8] Cleavage of R117-V118 appears to be an important step in the degradation of trypsin;[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.[7]

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

Uncertainty in time (D_{REG} and D_{SAS}) can be neglected because it would be small in relative terms. In D_{ADS}, uncertainty for the x-axis (equilibrium concentration of [T]) should be comparable to the uncertainty in the y-axis in D_{REG} and D_{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_{SAS} includes a change in pH from 2-3 in the stock solution to 8 at $t = 0$. If their equilibrium constant K_{Eq} depends on pH, A and B would not be in equilibrium at pH=8. Lacking this information, we assume that $K_{\mathrm{Eq}}^{\mathrm{pH=2-3}} = K_{\mathrm{Eq}}^{\mathrm{pH=8}}$.

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 trpysin 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. $[\mathrm{T}]_0 = [\mathrm{A}]_0$. 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

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}, \tag{S32}$$

where:

D = experimental data

θ = model parameters

$p(x|y)$= conditional probability of x given y

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

$$p(D|\theta) = \prod_i p_i(D_i|\theta_i). \tag{S33}$$

θ_i is a subset of the overall parameter space θ. 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:

$$p(\theta|D_i) = \frac{p_i(D_i|\theta) \prod_{j \neq i} p(\theta|D_j)}{p(D_i)} \tag{S34}$$

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 σ, using a normal distribution might lead to uncertainty intervals and posterior predictions of negative concentrations which are unphysical.

$$[\mathrm{X}]_t^{\mathrm{exp}} \sim N([\mathrm{X}]_t^{\mathrm{model}}, \sigma) \tag{S35}$$

where:

$[\mathrm{X}]_t^{\mathrm{exp}}$ = experimental concentration of X at time t

$[\mathrm{X}]_t^{\mathrm{model}}$ = modeled concentration of X at time t

σ = standard deviation

$$[\mathrm{X}]_t^{\mathrm{model}} = [\mathrm{X}]_0 + \int_0^t \frac{d[\mathrm{X}]}{dt} dt \tag{S36}$$

$\frac{d[\mathrm{X}]}{dt}$ depends on the kinetic model and its parameters. $[\mathrm{X}]^{\mathrm{model}} = [\mathrm{T}] + [\mathrm{TS}] = [\mathrm{A}] + [\mathrm{B}] +$

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

S6.1 Influence of σ

σ can be a constant or a nuisance parameter. JW1981 do not provide any error estimate, i.e. σ is unknown and treated as a nuisance parameter according to the `emcee` tutorial. [2]

$$\sigma = \sqrt{\sigma_{fix}^2 + f^2 y_{model}^2} \tag{S37}$$

where:

σ_{fix} = underestimated standard deviation

f = fractional deviation

Both σ_{fix} and f are inferred from the data. σ_{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, σ_i should equal σ_j. For JW1981, we assume that the experimental uncertainty of trypsin concentrations with or without silica surfaces is identical, hence $\sigma_{D_{\mathrm{SAS}}} = \sigma_{D_{\mathrm{REG}}}$. The experimental uncertainty $\sigma_{D_{\mathrm{ADS}}}$ 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] `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:

$$\mathrm{X} \xrightarrow{k_1} \mathrm{Y} \xrightarrow{k_2} \mathrm{Z} \tag{S38}$$

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.

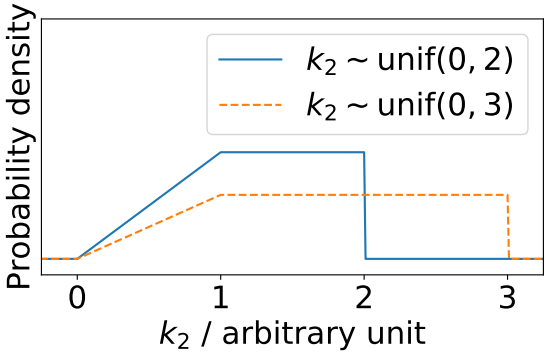


Figure S5: $p(\theta|D)$ for different uniform priors.

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. [3] We added two orders of magnitude to the lower and upper boundary as safety margin. Moreoever, our system describes the binding of enzyme molecules to binding sites, not to colloidal particles (CPs). Every such particle has $n_\mathrm{S/CP}$ binding sites on its surface. In order to yield accurate results, parameters should satisfy

$$k_\mathrm{a,model}[\mathrm{T}][\mathrm{S}] = k_\mathrm{a,real}[\mathrm{T}][\mathrm{CP}]. \tag{S39}$$

Since

$$[\mathrm{S}] = [\mathrm{CP}] \cdot n_\mathrm{S/CP} \tag{S40}$$

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

$$k_\mathrm{a,real} = k_\mathrm{a,model} \cdot n_\mathrm{S/CP}. \tag{S41}$$

Kinetically perfect enzymes have specificity constants $\frac{k_\mathrm{cat}}{K_m}$ (merged in our model to rate constant of the bulk reaction (k_BR)) of up to $10^8 - 10^9$ $\mathrm{M^{-1}\,s^{-1}}$, so we chose 10^9 $\mathrm{M^{-1}\,s^{-1}}$ as upper limit for k_BR.[12] The lower limit can be derived from the integrated rate law of a second-order reaction

$$\frac{1}{[\mathrm{T}]} = k_\mathrm{BR} t + \frac{1}{[\mathrm{T}]_0} \tag{S42}$$

and the data. Solving Eq. S42 for k_BR yields

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

We know from D_REG that for $t = 20000$ s, $[\mathrm{T}]_0 = 2.2$ µM, $\frac{[\mathrm{T}]}{[\mathrm{T}]_0} \ll 98\%$. After rounding we set

[3] `https://www.sprpages.nl/kinetics/association`; accessed at 15:48, August 2nd, 2019

the lower end of the uniform prior to $k_{\mathrm{BR}} > 10^{-0.5}\ \mathrm{M}^{-1}\,\mathrm{s}^{-1}$.

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

$$k'_{ER} = -t^{-1}\left(\log\frac{[\mathrm{T}]_0}{[\mathrm{T}]} + K_{\mathrm{a}}^{-1}\left(\frac{1}{[\mathrm{T}]_0} - \frac{1}{[\mathrm{T}]}\right)\right) \tag{S44}$$

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

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

$$\begin{aligned} k_{\mathrm{AB}} &\sim \mathrm{unif}(10^{-8}\ \mathrm{s}^{-1}, 10^{-2}\ \mathrm{s}^{-1}) \\ K_{\mathrm{Eq}} = \frac{k_{\mathrm{AB}}}{k_{\mathrm{BA}}} &\sim \mathrm{unif}(10^{-8}, 10^{0}) \end{aligned} \tag{S45}$$

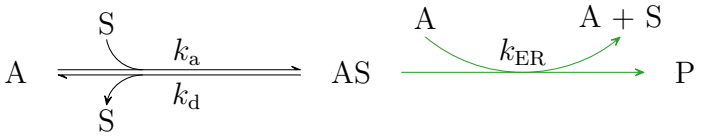
Priors for nuisance parameters are shown in Table S1.

Table S1: Priors for nuisance parameters

parameter	lower limit	upper limit	unit
σ_{kin}	$([\mathrm{T}] + [\mathrm{TS}])^{\mathrm{exp}}_{\mathrm{min}} \cdot 10^{-3}$	$([\mathrm{T}] + [\mathrm{TS}])^{\mathrm{exp}}_{\mathrm{max}}$	M
f_{kin}	10^{-4}	1	
σ_{ads}	$[\mathrm{TS}]^{\mathrm{exp}}_{\mathrm{min}} \cdot 10^{-3}$	$[\mathrm{TS}]^{\mathrm{exp}}_{\mathrm{max}}$	M
f_{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,[13] toxicokinetics[14] and system biology,[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 (θ_{true}) for the reactions shown in Scheme S3, aiming to reobtain θ via Bayesian inference. Data on kinetics (D_{KIN}) display the concentration of a protease A. Equilibrium adsorption (D_{EQ}) data track the adsorbed amount of A on a surface compared to the remaining unbound concentration. Time-dependent adsorption (D_{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_{KIN}. In both D_{EQ} and D_{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_{EQ} in BDI tightly restrains K_a and $[S]_0$. D_{TD} restrain $[S]_0$, the association rate (k_a) and to some extent

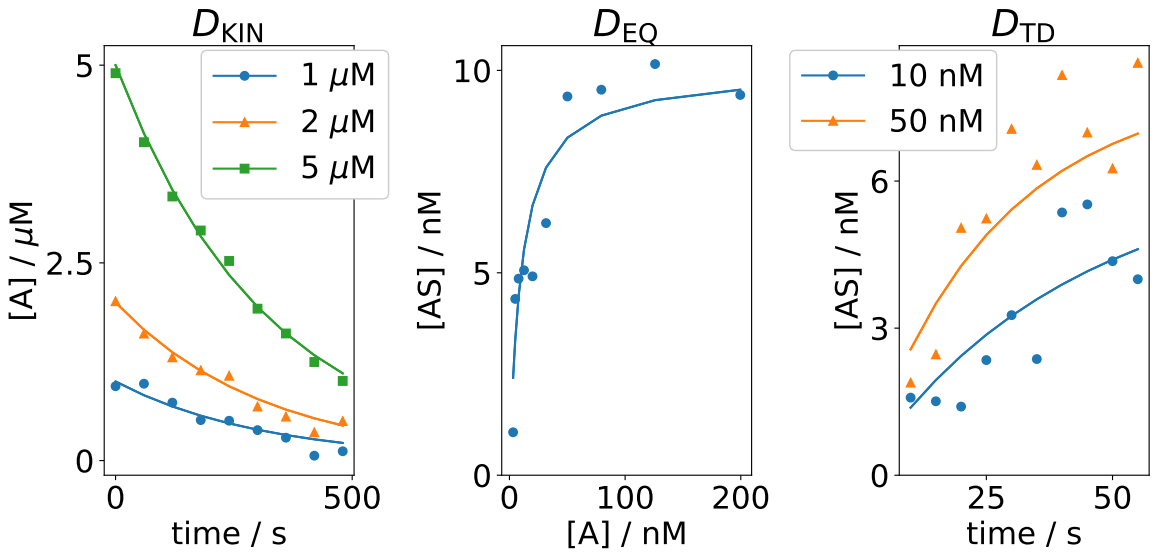


Figure S6: Three synthetic data sets based on θ_{true}. Lines correspond to the noise-free values generated from the models based on θ_{true}, markers include added noise. In D_{TD}, two starting concentrations of A ($[\mathrm{A}]_0$) were studied.

the dissociation rate (k_{d}). Combining all three data sets, the accessible parameter space shrinks and $k'_{ER} = k_{\mathrm{ER}}[\mathrm{S}]_0$ (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 $p(\theta_{\mathrm{true}}|D)$. This is a sign of (weak) overfitting: since the noise is random, no set of parameters can be consistently better than θ_{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

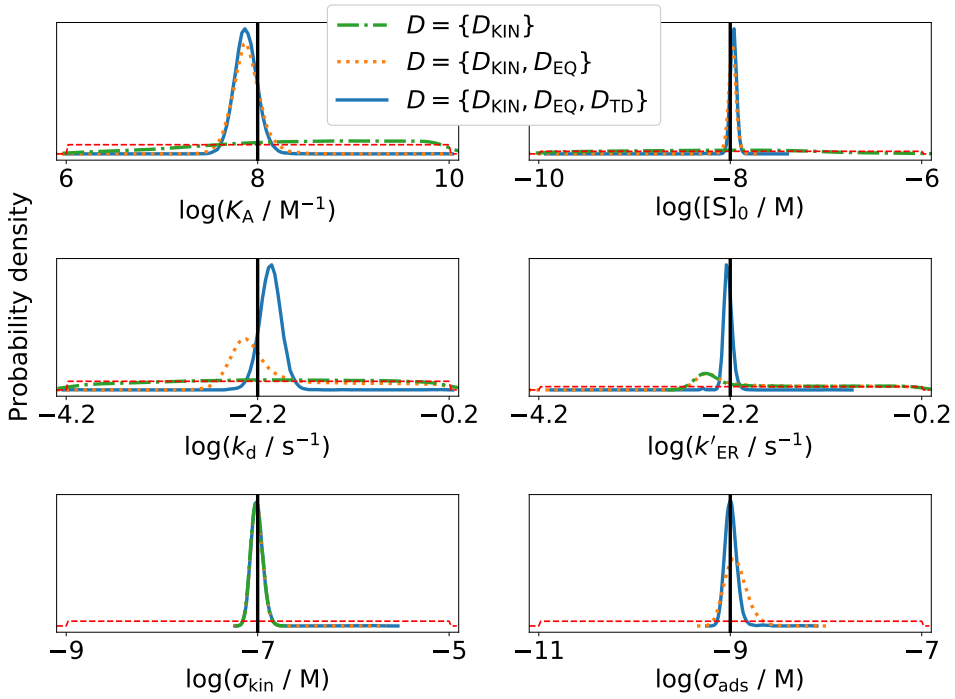


Figure S7: Marginalized posterior distributions $p(\theta|D)$ for $D = \{D_{\mathrm{KIN}}\}$ (green), $D = \{D_{\mathrm{KIN}}, D_{\mathrm{EQ}}\}$ (orange), $D = \{D_{\mathrm{KIN}}, D_{\mathrm{EQ}}, D_{\mathrm{TD}}\}$ (blue). True underlying parameter values θ_{true} are shown as black lines, priors are shown as red dashed lines. Refer to Section S6.2.2 for details why the prior is defined for $k'_{ER} = k_{\mathrm{ER}}[\mathrm{S}]_0$.

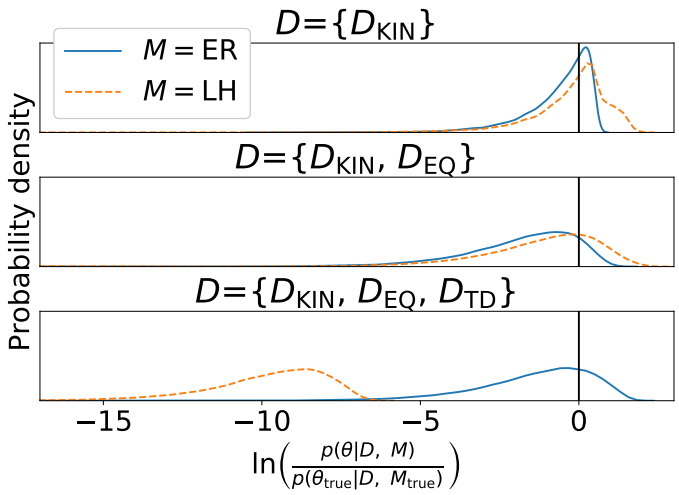


Figure S8: The maximum log posterior for ER and LH depends on the inclusion of different data sets.

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.; Patzscly, 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 α- and β-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.

3.3 Classifier uncertainty: evidence, potential impact, and probabilistic treatment

This section is based on the following publication:

[38] Tötsch, N & Hoffmann, D. (2020) Classifier uncertainty: evidence, potential impact, and probabilistic treatment. *arXiv* p. arxiv:2006.11105
`https://arxiv.org/abs/2006.11105`

Classifier uncertainty: evidence, potential impact, and probabilistic treatment

Niklas Tötsch[a,1] **and Daniel Hoffmann**[a]

[a] Bioinformatics and Computational Biophysics, Universität Duisburg-Essen, 45141 Essen, Germany

Classifiers are often tested on relatively small data sets, which should lead to uncertain performance metrics. Nevertheless, these metrics are usually taken at face value. We present an approach to quantify the uncertainty of classification performance metrics, based on a probability model of the confusion matrix. Application of our approach to classifiers from the scientific literature and a classification competition shows that uncertainties can be surprisingly large and limit performance evaluation. In fact, some published classifiers are likely to be misleading. The application of our approach is simple and requires only the confusion matrix. It is agnostic of the underlying classifier. Our method can also be used for the estimation of sample sizes that achieve a desired precision of a performance metric.

Classification | Machine Learning | Uncertainty | Bayesian Modeling | Reproducibility

Classifiers are ubiquitous in science and every aspect of life. They can be based on experiments, simulations, mathematical models or even expert judgement. The recent rise of machine learning has further increased their importance. But machine learning practitioners are by far not the only ones who should be concerned by the quality of classifiers. Classifiers are often used to make decisions with far-reaching consequences. In medicine, a therapy might be chosen based on a prediction of treatment outcome. In court, a defendant might be considered guilty or not based on forensic tests. Therefore, it is crucial to assess how well classifiers work.

In a binary classification task, results are presented in a 2×2 confusion matrix (CM), comprising the numbers of true positive (TP), false negative (FN), true negative (TN) and false positive (FP) predictions.

$$\mathrm{CM} = \begin{bmatrix} \mathrm{TP} & \mathrm{FN} \\ \mathrm{FP} & \mathrm{TN} \end{bmatrix} \quad [1]$$

CM contains all necessary information to determine metrics which are used to evaluate the performance of a classifier. Popular examples are accuracy (ACC), true positive rate (TPR), and true negative rate (TNR)

$$\mathrm{ACC} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{FN}+\mathrm{FP}+\mathrm{TN}} \quad [2]$$

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \quad [3]$$

$$\mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}} \quad [4]$$

These are given as precise numbers, irrespective of the sample sizes (Ns) used for their calculation in performance tests. This is problematic especially in fields such as biology or medicine, where data collection is often expensive, tedious, or limited by ethical concerns, leading often to small Ns. In this study we demonstrate that in those cases the uncertainty of the CM entries cannot be neglected, which in turn makes all performance metrics derived from the CM uncertain, too. In the light of the ongoing replication crisis (1), it is plausible that negligence of the metric uncertainty impedes reproducible classification experiments.

There is a lack of awareness of this problem, especially outside the machine learning community. One often encounters discussions of classifier performance lacking any statistical analysis of the validity in the literature. If there is a statistical analysis it usually relies on frequentist methods such as confidence intervals for the metrics or null hypothesis significance testing (NHST) to determine if a classifier is truly better than random guessing. NHST "must be viewed as approximate, heuristic tests, rather than as rigorously correct statistical methods" (2).

Bayesian methods can be valuable alternatives. (3) To properly account for the uncertainty, we have to replace the point estimates in the CM and all dependent performance metrics by probability distributions. Correct and incorrect classifications are outcomes of a Binomial experiment. (4) Therefore, Brodersen et al. model ACC with a beta-binomial distribution (BBD)

$$\mathrm{ACC} \sim \mathrm{Beta}(\mathrm{TP}+\mathrm{TN}+1, \mathrm{FP}+\mathrm{FN}+1). \quad [5]$$

Some of the more complex metrics, such as balanced accuracy, can be described by combining two BBDs. (4)

Caelen presented a Bayesian interpretation of the CM. (5) This elegant approach, based on a single Dirichlet-multinomial distribution, allows to replace the count data of the confusion matrix with distributions which account for the uncertainty.

$$\mathrm{CM} \sim \mathrm{Mult}(\theta, N) \quad [6]$$

$$\theta \sim \mathrm{Dirichlet}((1,1,1,1)) \quad [7]$$

where $\theta = [\theta_{\mathrm{TP}}, \theta_{\mathrm{FN}}, \theta_{\mathrm{TN}}, \theta_{\mathrm{FP}}]$ is the confusion probability matrix which represents the probabilities to draw each entry of the CM. The major advantage of Caelen's approach over the one presented by Brodersen lies in a complete description of the CM. From there, all metrics can be computed directly, even those that cannot simply be described as BBD.

Caelen calculates metric distributions from confusion matrices that are sampled according to Equation 6. Here, we demonstrate that this approach is flawed and derive a correct model. Whereas previous studies focused on the statistical methods, we prove that classifier performance in many peer-reviewed publications is highly uncertain. We studied a variety

N.T. and D.H. designed research; N.T. performed research and analyzed data; N.T. and D.H. wrote the paper.

The authors declare no conflict of interest.

[1] To whom correspondence should be addressed. E-mail: niklas.toetsch@uni-due.de

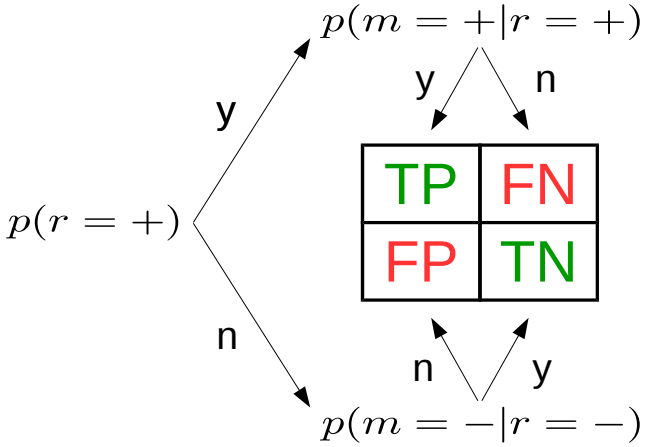


Fig. 1. Three beta-binomial distributions $p(\cdot)$ – prevalence (left), true positive rate (top), true negative rate (bottom) – define the confusion matrix. Based on them, all entries of the CM can be expressed as distributions with explicit uncertainty due to limited sample size.

of classifiers from the chemical, biological and medicinal literature and found cases where it is not clear if the classifier is better than random guessing. Additionally, we investigate metric uncertainty in a Kaggle machine learning competition where sample size is relatively large but a precise estimate of the metrics is required. In order to help non-statisticians to deal with these problems in the future, we derive a rule for sample size determination and offer a free, simple to use webtool to determine metric uncertainty.

1. Methods

A. Model. The confusion probability matrix (θ), that is the probabilities to generate entries of a confusion matrix, can be derived if prevalence (ϕ), TPR and TNR are known. (6)

$$\theta_{\mathrm{TP}} = \mathrm{TPR} \cdot \phi \quad [8]$$
$$\theta_{\mathrm{FN}} = (1 - \mathrm{TPR}) \cdot \phi \quad [9]$$
$$\theta_{\mathrm{TN}} = \mathrm{TNR} \cdot (1 - \phi) \quad [10]$$
$$\theta_{\mathrm{FP}} = (1 - \mathrm{TNR}) \cdot (1 - \phi) \quad [11]$$

The idea that these metrics can also be inferred from data, propagating the uncertainty, is the starting point of the present study. Using three BBDs, one for each of ϕ, TPR and TNR, we can express all entries of the CM (Figure 1). Since ϕ, TPR and TNR are distributions, the entries of θ $[\theta_{\mathrm{TP}}, \theta_{\mathrm{FN}}, \theta_{\mathrm{TN}}, \theta_{\mathrm{FP}}]$ are too. Based on θ we calculate all other metrics of interest.

For the following Bayesian treatment we use the Laplace prior, $\mathrm{Beta}(\alpha = 1, \beta = 1)$, for ϕ, TPR and TNR because its uniform distribution introduces no bias, which makes it suitable for any classification problem. It is noteworthy that a flat prior on ϕ, TPR and TNR leads to non-flat priors on other metrics (Section S1). We discuss two additional objective priors in the supplementary material. If additional knowledge is available, based e.g. on the experimental setup of the classifier, it should be incorporated in the prior. Here, we refrain from using informative priors to keep the method generally applicable.

Our approach is quite similar to Caelen's but has distinct advantages. First, ϕ, TPR and TNR are common metrics; thus prior selection is easier. Second, our model clearly distinguishes data intrinsic ϕ from the classifier intrinsic measures TPR and TNR. Consequently, our approach allows to "exchange" ϕ. This is useful if the prevalence of the test set differs from the prevalence of the population the classifier will be applied to in production. Such a scenario is common in medical tests where ϕ is very low in the general population. To increase the sample size of positive cases in the test set without inflating the number of negative ones, ϕ differs from the general population. Using a Dirichlet-multinomial distribution, it is not straightforward to evaluate a classifier for a different ϕ. If the data set was designed to contain a specified fraction of positive and negative instances, ϕ is known exactly (Section S2). This scenario is easy to implement in our model but not in Caelen's.

Depending on the context, ϕ may have two meanings. If one is interested in a population, ϕ describes how common fulfilment of the positive criterion is. For an individual, e.g. a patient, ϕ can be considered the prior. If additional information was available for this subject, such as results of previous tests, ϕ would differ from the prevalence in the general population. This prior can be updated with TPR and TNR, representing the likelihood, to yield the posterior for the individual.

B. Measuring true rather than empirical uncertainty. Bayesian models allow posterior predictions. In our case, posterior predictions would be synthetic confusion matrices V, which can be generated from a multinomial distribution (Equation 6).

This approach is equivalent to a combination of two/three binomial distributions as shown in Figure 1 but slightly more elegant for posterior predictions. Caelen samples many V to obtain metric distributions, which requires a choice of sample size N. Caelen uses the N of the original CM the parameters have been inferred from. This is not satisfying because in this way only the empirical distribution of the metrics for a given N is generated, not the true distribution of the metrics. Consider the example of CM = (TP, TN, FP, FN) = (1, 0, 0, 0), i.e. $N = 1$. We will consider this classifier's ACC. Caelen's approach leads to a discrete distribution of the accuracy allowing only 0 and 1 (Figure 2, top). There was one correct prediction in the original CM, therefore it is impossible that the accuracy is 0. In other words, the probability mass at ACC=0 should be strictly 0. If one is interested in the true continuous posterior distribution of a metric, one must calculate it from θ directly (Figure 2, bottom). We prove in Section S4 that Caelen's approach systematically overestimates the variance in metric distributions.

We still consider Caelen's way of calculating metrics extremely useful since it allows to tackle the problem of reproducibility. Generating synthetic V according to Equation 6 allows us to estimate what would happen if multiple researchers applied the same classifier to different data sets of size N and reported the corresponding CMs and metrics. Figure 2 shows that they might report completely different values of a metric if N is small. Under these circumstances, classification experiments are not reproducible.

C. Metric uncertainty equals credible interval length. If there is little data available, posterior distributions are broad. We define metric uncertainty (MU) as the length of the 95% highest posterior density interval ("credible interval"). There

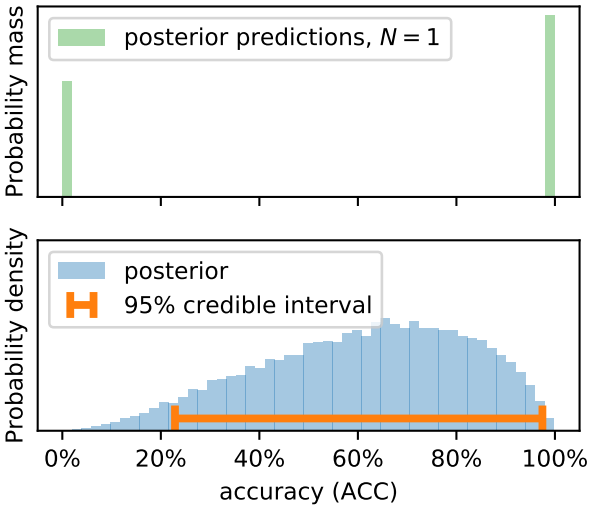


Fig. 2. Calculating accuracy (ACC) on posterior predictions of the confusion matrix yields a discrete distribution (top), representing expected observations of the metric at given sample size (N). Posterior distributions (bottom) of the metric must be calculated from the inferred entries of the confusion probability matrix (θ) as outlined in the text.

is a 95% likelihood that the metric is within this credible interval (bottom of Figure 2).

D. Implementation. Since the beta distribution is the conjugate prior of the binomial distribution, the posterior distribution can be derived analytically. There is no need for Markov chain Monte Carlo sampling. This is merely a convenience, our approach would work with any prior. To calculate metrics, we sampled 20000 data points. Splitting these data points into two arrays of equal length, we use `PyMC`'s implementation of the Gelman-Rubin diagnostics ($R_c < 1.01$) to verify that the posterior distribution is properly sampled. (7–9)

The implementation of our model in `Python` can be found on https://github.com/niklastoe/classifier_metric_uncertainty.

2. Results and Discussion

A. Classifier examples from the literature. To assess the uncertainty in classifier performance in the scientific literature, we searched `Google Images` for binary confusion matrices from peer reviewed publications in the area of chemistry, biology and medicine with less than 500 samples in the test set. We collected 24 classifiers; confusion matrices and the references to the publications are listed in Table S1. Publications are indexed with numbers. If more than one classifier is presented in one publication, a character is added. Some of these classifiers are based on statistical models of available data. Others are based on simulations. The majority of publications describe the development of a new experimental approach followed by a statistical model that transforms the experimental outcome into a classification. Classifiers come from diverse fields, e.g. chemical detection (adulterants in palm oil or cocaine, mycotoxins in cereals) or prediction of inhibitors of amyloid-aggregation or enzymes. The smallest sample size was 8, the largest 350.

While the resources invested in the development of these classifiers must have been considerable, their performance had not been thoroughly evaluated. Specifically, only for a single classifier the uncertainty had been quantified by calculating confidence intervals. In some of the literature examples, we also noted severe problems unrelated to small N. Due to usage of ACC for imbalanced data sets and mixing of train and test data sets for reported metrics, the performance of some classifiers was overrated. These problems have been addressed previously. (10) In this study, we evaluate classifiers on metrics which are invariant to class imbalance and rely exclusively on test data sets.

Our selection may not in all aspects be representative of published classifiers in any field. However, the negligence of metric uncertainty observed in this selection is not exceptional. Our choice of biology, chemistry, and medicine as scientific domain was based on our relative familiarity with those fields. While in this domain small sample sizes are common (due to costly data collection), this problem is probably not limited to this domain.

B. Metrics are broadly distributed. Typically, classifier metrics are reported as single numerical values (often to one or more decimals) without indication of uncertainty. However, the true MUs of classifiers in our collection are too large to be ignored (Figure 3). Often, MU is greater than 20 percentage points, sometimes exceeding 60 percentage points. In general, MU in all three observed metrics declines as N increases. The decrease is not monotonous because MU also depends on the value of the metric (Section S5).

The MUs we show in Figure 3 were obtained from θ. As mentioned above, metrics calculated from empirically observed confusion matrices of the same classifier would vary even more. Thus, if an independent lab tried to reproduce CM for, say, example 7a, with a much larger sample size, TNR values of 90% or 50% would not be surprising, although the value given in the paper is 75%.

It is possible that we underrate some classifiers. If a metric should have a more informative prior than the Laplace prior we used, e.g. due to previous experience or convincing theoretical foundations, the posterior could also be more narrowly defined.

B.1. Metric uncertainty limits confidence in high-stakes application of classifiers. In the following, we discuss in greater detail MU for one classifier where the consequences of misclassification are dramatic and understandable to non-experts. Classifier 7a is a new method to predict cocaine purity based on a "simple, rapid and non-destructive" experiment followed by mathematical analysis. The authors stress the importance of such a method for forensic experts and criminal investigators. Predictions are compared to a destructive and more elaborate experimental reference. Prosecutors in countries such as Spain may consider purity as evidence of the intent to traffic a drug, presumably resulting in more severe punishments.* Consequently, a FP would result in a wrongful charge or conviction causing severe stress and eventually imprisonment for the accused. A FN on the other hand might lead to an inadequately mild sentence. Moreover, one could also consider the scenario of drug checking. In some cities, such as Zurich, Switzerland, social services offer to analyze drugs to prevent harm from substance abuse

*http://www.emcdda.europa.eu/system/files/publications/3573/Trafficking-penalties.pdf; accessed December 3rd, 2019

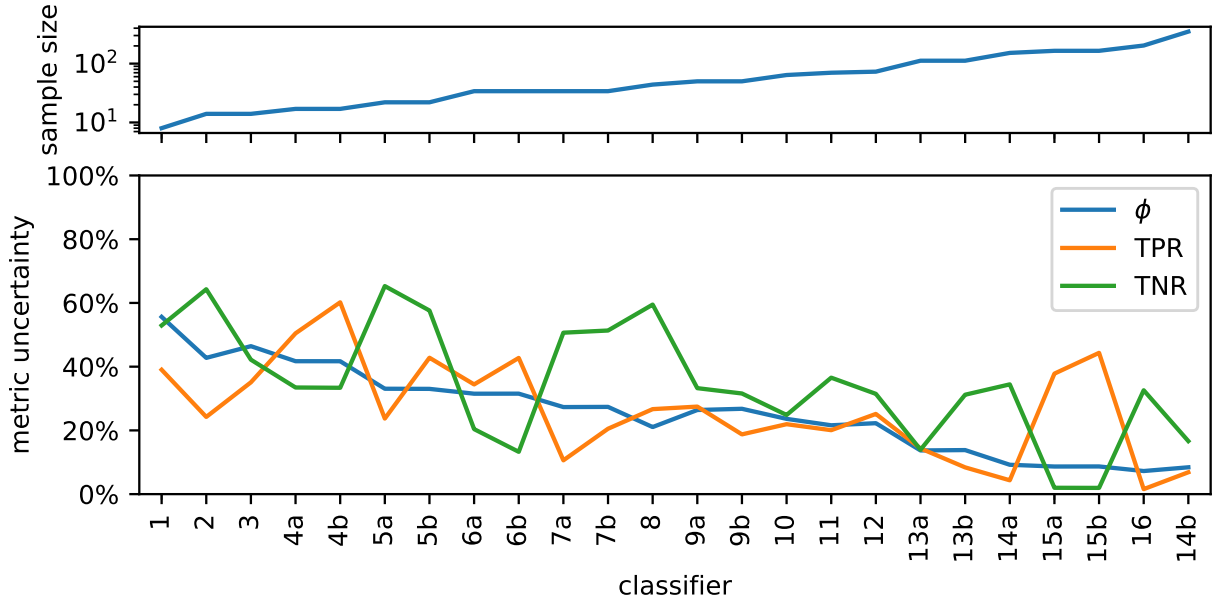


Fig. 3. Metric uncertainty (MU) for prevalence (ϕ), true positive rate (TPR), and true negative rate (TNR) of literature examples. Classifiers are sorted by sample size (top). MU is large but decreases with sample size. Since MU is determined by the length of the 95% highest posterior density interval, the theoretical upper limit is 95% (in which case little is known about the metric). If MU was 0%, the corresponding metric would be known at infinite precision.

due to unexpectedly high purity or toxic cutting agents.[†] In this context, a FN could lead to an overdose due to the underestimated purity.

The confusion matrix in Figure 4a is transcribed from the original publication. We do not know whether their method was used for drug checking or in court (at least the authors received the samples from the local police department). If it was, could it be trusted by a forensic expert, judge, or member of the jury? The posterior distribution of the TPR (Figure 4b) answers this question probabilistically. The point estimate from CM would be TPR=100% but due to small N, the uncertainty is large. The credible interval spans from 89% to almost 100% although not a single FN has been observed in the test set.

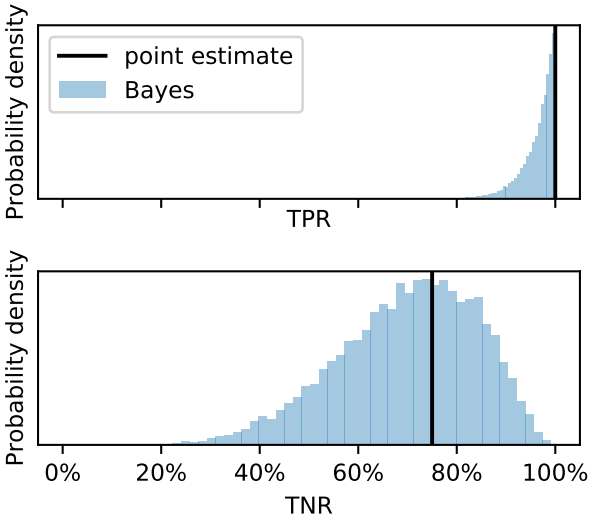
Now consider TNR. Since there are only eight low purity cocaine samples, the uncertainty is much larger. While the point estimate would be TNR=75%, the credible interval is 43%-95%. It is possible, although unlikely, that the classifier would generate more FP than TN. This would translate into more wrongful convictions than correct acquittals for possessing cocaine with high purity if this method was used as main evidence in court.

Our approach would hopefully lead to more cautious use of little tested classifiers. Imagine two scenarios. In the first, a judge is told that the forensic method has a TPR of 100% and a TNR of 75%. In the second, she is told that it has an estimated TPR of 89-100% and an estimated TNR of 43-95%. In the latter, the judge would be more hesitant to base her verdict on the classifier.

We do not know if ϕ in the test set is representative of the prevalence of drug samples in criminal cases. Therefore, we cannot reasonably estimate the distribution of probabilities of wrongfully harsh/lax sentences. For a meaningful assessment

	r=high	r=low
m=high	26	2
m=low	0	6

(a) Confusion matrix, r stands for reference, m for model



(b) Posterior distributions for true positive rate (TPR) (top) and true negative rate (TNR) (bottom)

Fig. 4. Metric uncertainty for cocaine purity classifier 7a

[†] https://www.saferparty.ch/worum-gehts.html; accessed on June 9th, 2020 at 3:42 pm

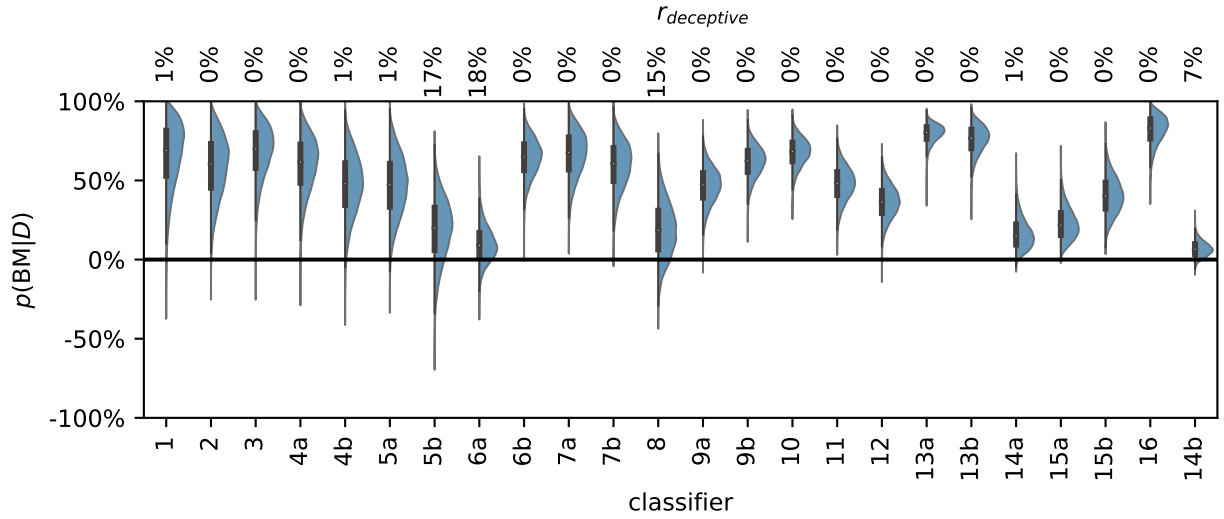


Fig. 5. Posterior distributions of bookmaker informedness (BM) are broad due to small test sets in the literature examples. Classifiers are sorted by ascending sample size as in Figure 3. Some classifiers have considerable posterior density in the negative region; these classifiers could be misinformative. Percentages along top margin are $r_{\text{deceptive}}$ values (Equation 14), the probability that a classifier is worse than random guessing.

of evidence, both ϕ and MU should be taken into account. Our approach facilitates such an analysis.

C. Some published classifiers might be deceptive. As classification problems vary greatly so does the relevance of different metrics, depending on whether FN or FP are more or less acceptable. Often, classifier development requires a tradeoff between FN or FP. In this respect, bookmaker informedness (BM) is of interest because it combines both in a single metric without weighting and measures the probability of an informed prediction. (11)

$$\text{BM} = \text{TPR} + \text{TNR} - 100\% \quad [12]$$

If BM=100%, prediction is perfect and the classifier is fully informed. BM=0% means that the classifier is no better than random guessing and BM=-100% shows total disagreement, i.e. the predictor is wrong every single time. Figure 5 shows the posterior distributions of BM for the collected examples from literature. Due to small N, they are broad. Therefore, it is uncertain how much better the classifiers are compared to random guessing. Several classifiers have considerable probability density in the negative region, i.e. it is possible that they are weakly deceptive.

We define the probabilities that a given classifier is informative or deceptive

$$r_{\text{informative}} = \int_{0\%}^{100\%} p(\text{BM}|D)d\theta \quad [13]$$

$$r_{\text{deceptive}} = \int_{-100\%}^{0\%} p(\text{BM}|D)d\theta. \quad [14]$$

We determined $r_{\text{deceptive}}$ for all literature examples (Figure 5, top). Four classifiers have a considerable chance to be deceptive. We note that three of them were published alongside alternative classifiers that the respective authors considered preferable (5b, 6a, 14b). The probability that the classifier 8 is deceptive is approximately 15% so we recommend to reevaluate it with a larger test set.

The split of the BM posterior into $r_{\text{informative}}$ and $r_{\text{deceptive}}$ in Equations 13 and 14 is a coarse-graining device to ease conversation. A classifier with a very low absolute BM is neither informative nor deceptive but uninformative.

For finite N, $r_{\text{deceptive}}$ will be always greater than zero. What value of $r_{\text{deceptive}}$ can be tolerated will of course depend on the application scenario, and should be carefully considered by developers and users of classifiers.

D. Large N, small difference in performance in metaanalysis of classifiers in machine learning. Our approach can also be used for meta-analyses of classifier ensembles, an application that is of considerable interest in machine learning. (2, 3, 12) Kaggle, a popular online community for machine learning challenges, provides a suitable environment for such meta-analyses. On Kaggle, participants build classifiers and submit their results online to be evaluated and compared to those of others. The best results are rewarded with cash prizes. The metric for evaluation depends on the individual challenge. Often, the competition is fierce and submitted results close, e.g. accuracy sometimes differs by less than one per mille. With hundreds to tens of thousands of data points, test sets tend to be larger than in our literature collection above, but are still finite. Classifier metrics therefore retain some uncertainty, and statistical flukes could produce apparent differences in classifier performances that decide a competition.

We studied the Recursion Cellular Image Classification competition in greater detail.[‡] Participants are tasked to properly classify biological signals in cellular images, disentangling them from experimental noise. Submissions were ranked based on multiclass accuracy. Micro-averaged multiclass accuracy can be modeled according to Equation 5. We evaluated private leaderboards, i.e. rankings provided by Kaggle with informa-

[‡] https://www.kaggle.com/c/recursion-cellular-image-classification/overview; accessed on January 31st, 2020 at 9:25 am

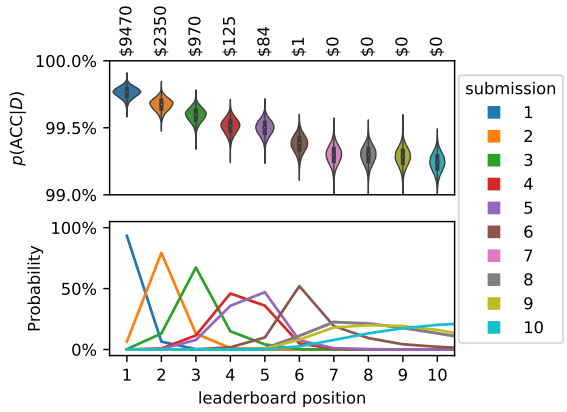


Fig. 6. Accuracy (ACC) posterior distribution for top ten submissions on Kaggle leaderboard (top). Distributions are narrow but the classifiers perform similarly. Therefore, after consideration of the uncertainty in ACC, the leaderboard positions of the submissions are uncertain (bottom). If the cash prizes were awarded based on the probabilistic leaderboard, submissions outside of the top three would receive money (annotation). These estimates, too, are uncertain by a few percentage points.

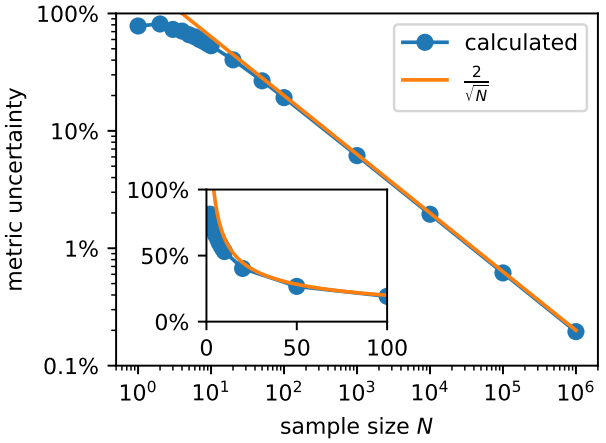


Fig. 7. Sample size determines metric uncertainty (defined by the length of the 95% highest posterior density interval) for any metric whose distribution follows a BBD. Statistical power is 95%. The inset shows the same data on a non-logarithmic scale.

tion on the participants and accuracies of their classifiers. These private leaderboards were also used to award prizes. Kaggle did not publish the exact size of the private test set but the overall test set contains 19899 images and the private leaderboards were calculated on approximately 76% of it so we assumed N=15123. Based on N and the published point estimates of ACC we could calculate TP+TN and FP+FN for every submitted classifier and compute a posterior distribution for ACC according to Equation 5 (Figure 6, top).

These posterior distributions overlap. Using a Monte Carlo approach, we generated synthetic leaderboards from samples of the posterior distributions. Counting how often every submission occurred at any leaderboard position yielded a probabilistic leaderboard (Figure 6, bottom). We observed that the winning submission has a 93% chance of being truly better than any other submission. For leaderboard position 4 and worse, rank uncertainty becomes considerable and ranking validity is limited by the sample size.

At the end of this competition, the top 3 submission were awarded $10.000, $2.000 and $1.000, respectively. This implies that it is certain that the submissions listed in the top 3 positions are indeed the best classifiers. As we have demonstrated, it is not certain which submissions are the best. If one would weigh the awarded prizes based on the probability of a submission to be in each rank, other participants would have been awarded small cash prizes (Figure 6, top annotation).

Our approach is complementary to the Bayesian Plackett-Luce model, which considers multiple rankings for individual problems. (12) That model is agnostic about the performance metric since it is based only on the leaderboard position in every scenario. Consequently, it neglects the magnitude of the performance difference. Our approach on the other hand requires a generative model for the performance metric but works for individual problems and quantifies the performance gap between classifiers.

E. Sample size determination. Since uncertainty in any commonly used metric decreases with increasing sample size N, we can employ our approach of uncertainty quantification also to determine in advance values of N so that a classifier fulfills predefined MU criteria.

For those metrics which can be described as BBD (Equation 5), such as ACC, TPR, TNR and ϕ, we tested N values spanning six orders of magnitude (Figure 7), following Kruschke's protocol for sample size determination. (13) The shown results were obtained for a generating mode ω=0.8 and concentration $k = 10$. We found that different ω yielded almost indistinguishable results at low k.

The probability to achieve a MU more narrow than the given width in an empirical study, i.e. statistical power, is 95%. The interpretation is as follows: If N=100, the likelihood that $MU \leq 19$ percentage points is 95%. In order to decrease MU further, N must be increased substantially.

Based on the standard deviation of a beta distribution and the central limit theorem we derive

$$MU \lessapprox \frac{2}{\sqrt{N}} \qquad [15]$$

for $N > 20$ in Section S5. It yields the correct order of magnitude which tells us if a classification study is feasible at the desired level of MU. This general rule ignores prior knowledge about the classifier. The posterior of the metric derived from exploratory classification experiments should be considered.

We found several papers presenting metrics with one or even two decimals. Classifier evaluations should be considered like any other experiment, and only significant digits should be given in their discussion. Equation 15 predicts that metric uncertainty would only drop below 0.1%, which is necessary to present a metric with a decimal, if the test data set included several million data points. Curating such a large test set is out of the question for the publications in our examples. On Kaggle leaderboards, ACC is presented as percentage with three decimals. Reducing metric uncertainty below 0.001% would require tens of billions of data points.

3. Conclusions

In this work, we have presented a Bayesian model that quantifies the metric uncertainty of classifiers due to finite test sets.

It is completely agnostic about the underlying classifier. Unlike previous work, our method cleanly separates data intrinsic ϕ from classifier intrinsic TPR and TNR, which facilitates transfer to different data sets. Nevertheless, our approach allows to evaluate metric uncertainty of all metrics that are based on the CM.

Our study of published examples suggests that MU is a neglected problem in classifier development. We found classifier metrics that were typically highly uncertain, often by tens of percentage points. The respective articles do not address this uncertainty, regularly presenting insignificant figures. Therefore, their audience is unintentionally mislead into believing that classifier metrics are known precisely although this is clearly not the case.

We could show that some classifiers carry a non-negligible risk of being deceptive. Moreover, empirical uncertainties, determined by repeating a classification experiment, would be even larger than the true uncertainty of a metric due to small N. Thus, many published classification metric point estimates are unlikely to be reproducible.

Poorly understood classifiers potentially harm individuals and society. Our example on cocaine purity analysis has shown that the number of miscarriages of justice due to an insufficiently tested classifier could be alarmingly high. Similarly, the likelihood of misdiagnoses and subsequent wrongfully administered therapies based on a medical classifier remain obscure unless we account for sample size. In basic science, uncertain classifiers can misguide further research and thus waste resources. During the identification of molecules with therapeutic potential, a poor classifier would discard the most promising ones or lead the researchers to a dead-end. Since time and funding are finite, this would decrease progress resulting in economic as well as medical damages.

The example of the Kaggle challenge shed light on the problem of uncertain performance in classifier meta-analysis. There, sample size is usually large but performance differences are minute. Consequently, classifier or algorithm rankings are uncertain.

We can interpret the frequent failure to account for metric uncertainty in classification as another facet of the current replication crisis, one root cause of which is neglect of uncertainty. (14, 15) Classifier evaluation should be considered like any other experiment. It is obvious that a physical quantity cannot be measured exactly, and neither can a classifier metric. Thus, its uncertainty should be estimated and properly communicated.

For easy access to the method proposed here, we provide a free open-source software at https://github.com/niklastoe/classifier_metric_uncertainty. The software can be used without programming in an interactive web interface. The only required input is the confusion matrix, i.e. information that is usually available for published classifiers. The software then computes uncertainty for any of the commonly used classifier metrics. Moreover, sample sizes that are required to achieve a given exactness of a metric can be estimated according to Equation 15. We hope this contributes to more realistic expectations, more thoughtful allocation of resources and ultimately reliable performance assessments of classifiers.

Our approach can be extended to similar problems. Multiclass classification can be modeled by $c + 1$ multinomial distributions (where c is the number of classes), analogously to Figure 1. Another extension of our approach is the computation of error bars of the popular receiver operating characteristic (ROC) curve, which is basically a vector of CMs. It would be more difficult to use our approach to compute the uncertainty of the area under the ROC curve (AUC), another popular classifier metric. However, the AUC, too, will be uncertain for finite N. A further extension is the inclusion of classification scores in a distributional model (16), because the scores contain additional information that leads to a better understanding of MU.

Our approach only captures the uncertainty arising from finite N. Other sources of uncertainty such as over- or underfitting, data and publication bias etc. need to be considered separately. For instance, comparison of metric posterior distributions calculated separately for the training and test data could help to assess overfitting. Without such additional analyses, the posterior distributions obtained with our method are probably often too optimistic.

ACKNOWLEDGMENTS. We thank Paul Bürkner, Kai Horny, and Martin Theissen for fruitful discussion. Funding from Deutsche Forschungsgemeinschaft through project CRC1093/A7 is gratefully acknowledged.

1. Baker M (2016) Is there a reproducibility crisis? *Nature* 533(26):353–366.
2. Dietterich TG (1998) Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. *Neural Comput.* 10(7):1895–1923.
3. Benavoli A, Corani G, Demšar J, Zaffalon M (2017) Time for a change: A tutorial for comparing multiple classifiers through Bayesian analysis. *J. Mach. Learn. Res.* 18:1–36.
4. Brodersen KH, Ong CS, Stephan KE, Buhmann JM (2010) The balanced accuracy and its posterior distribution. *Proc. - Int. Conf. Pattern Recognit.* pp. 3121–3124.
5. Caelen O (2017) A Bayesian interpretation of the confusion matrix. *Ann. Math. Artif. Intell.* 81(3-4):429–450.
6. Kruschke JK (2015) Bayes' Rule in *Doing Bayesian Data Anal.* (Elsevier), pp. 99–120.
7. Gelman A, Rubin DB (1992) Inference from Iterative Simulation Using Multiple Sequences. *Stat. Sci.* 7(4):457–472.
8. Brooks SP, Gelman A (1998) General Methods for Monitoring Convergence of Iterative Simulations. *J. Comput. Graph. Stat.* 7(4):434–455.
9. Salvatier J, Wiecki TV, Fonnesbeck C (2016) Probabilistic programming in Python using PyMC3. *PeerJ Comput. Sci.* 2(4):e55.
10. Chicco D (2017) Ten quick tips for machine learning in computational biology. *BioData Min.* 10(1):1–17.
11. Powers DMW (2011) Evaluation: From Precision, Recall and F-Measure To Roc, Informedness, Markedness & Correlation. *J. Mach. Learn. Technol.* 2(1):37–63.
12. Calvo B, et al. (2019) Bayesian performance analysis for black-box optimization benchmarking in *Proc. Genet. Evol. Comput. Conf. Companion - GECCO '19.* (ACM Press, New York, New York, USA), pp. 1789–1797.
13. Kruschke JK (2015) *Goals, Power, and Sample Size.* pp. 359–398.
14. Gelman A, Carlin J (2017) Some Natural Solutions to the p -Value Communication Problem—and Why They Won't Work. *J. Am. Stat. Assoc.* 112(519):899–901.
15. Wasserstein RL, Schirm AL, Lazar NA (2019) Moving to a World Beyond "p < 0.05". *Am. Stat.* 73(sup1):1–19.
16. Brodersen KH, Ong CS, Stephany KE, Buhmann JM (2010) The binormal assumption on precision-recall curves. *Proc. - Int. Conf. Pattern Recognit.* (Section IV):4263–4266.

Supplementary Information for

Classifier uncertainty: evidence, potential impact, and probabilistic treatment

Niklas Tötsch, Daniel Hoffmann

Niklas Tötsch
E-mail: niklas.toetsch@uni-due.de

This PDF file includes:

Sections S1 to S5
Figs. S1 to S2
Table S1
References for SI reference citations

S1. Priors

The prior distribution can be interpreted as expression of previous knowledge, which in turn can be expressed in terms of previous observations. In this sense, the Laplace (or flat) prior is equivalent to two previous observations for each prevalence (ϕ), true positive rate (TPR) and true negative rate (TNR), which is usually a questionable assumption. Since sample size (N) is small in some of the examples discussed in this study, this assumption could have an impact on the posterior distribution. Nevertheless, we consider this prior to be the most suitable objective prior. Haldane's prior, $\mathrm{Beta}(\alpha = 0, \beta = 0)$, is not adequate since it yields an improper posterior if any entry of the confusion matrix (CM) is zero, which is often the case. Jeffreys prior, $\mathrm{Beta}(\alpha = 0.5, \beta = 0.5)$, does not have this problem but leads to implausible U-shaped priors for some metrics (Figure S2).

S2. Marginals of the confusion matrix

There are three scenarios for the marginals of the CM. In principle, the marginals of the columns and rows of the CM could both be fixed, which would mean that ϕ and the number of positive/negative predictions are known exactly beforehand. Fisher's exact test was designed to evaluate whether a binary classifier performs better than random guessing for this specific case. (1) It remains popular, yet the underlying assumption is usually violated. (2, 3)

A fixed ϕ and an unspecified marginal on the predicted labels is more common. For instance, in a controlled study, test sets may be curated to include 50% patients suffering from a disease and 50% healthy subjects in a control group. In this example there is no uncertainty in ϕ, but it is fixed at ϕ=0.5.

If ϕ in the test set is not deliberately chosen before the compilation, it must be determined from the data set. For small sample sizes, ϕ is uncertain like all other metrics. In the present study, we infer ϕ from the CM but our method also copes with fixed ϕ.

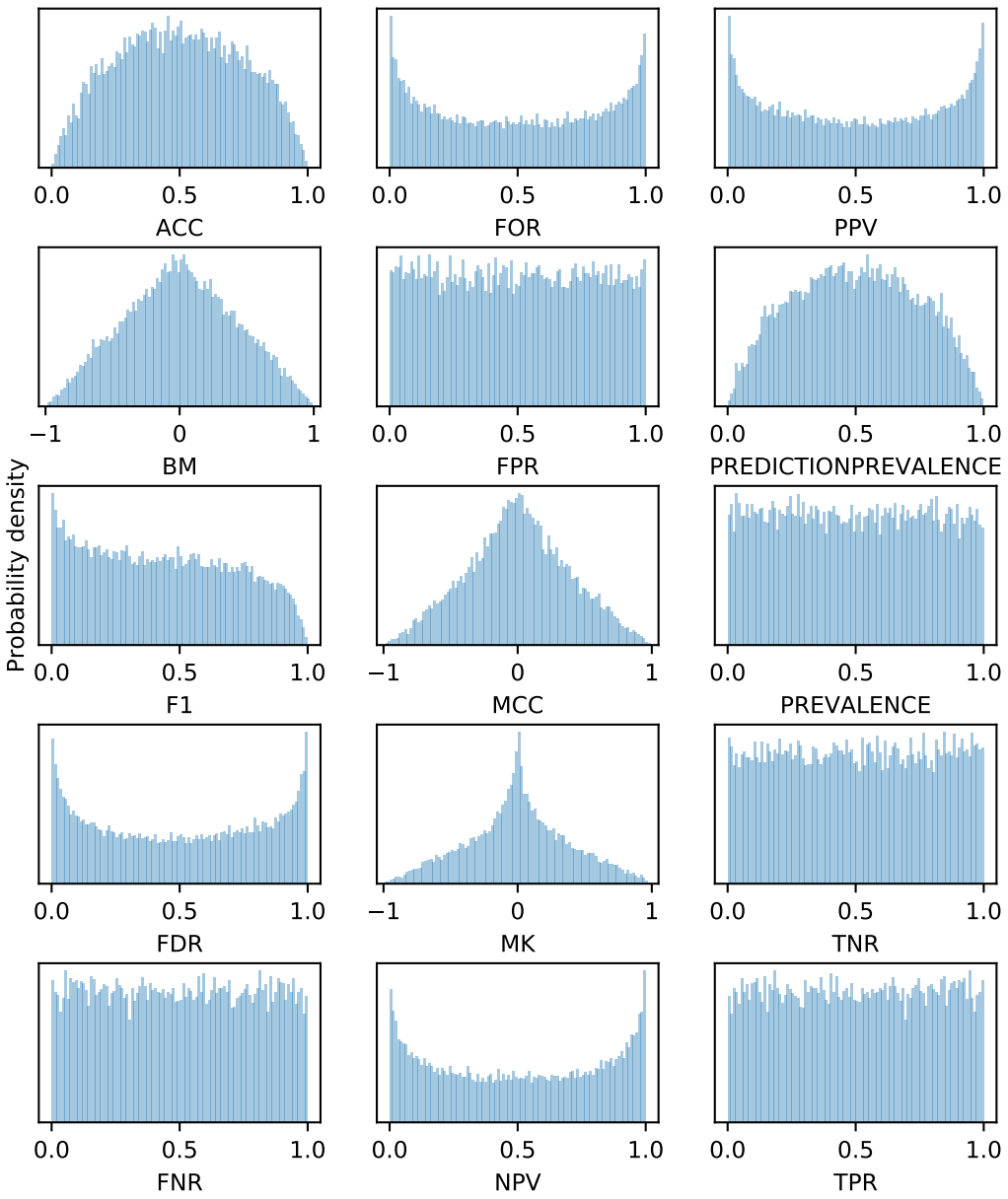


Fig. S1. Priors on the metrics if Laplace priors are used for ϕ, TPR, TNR

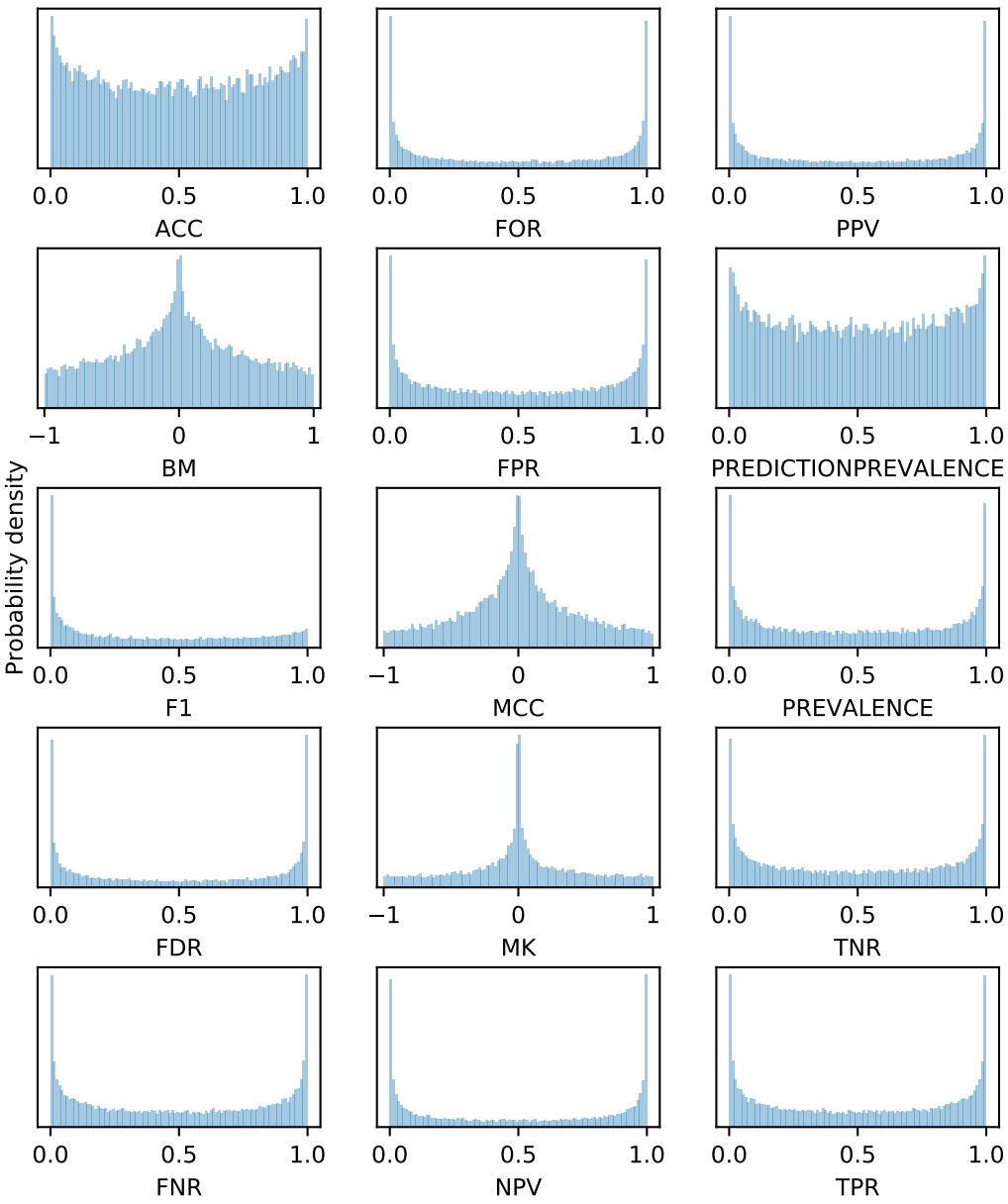


Fig. S2. Priors on the metrics if Jeffreys priors are used for ϕ, TPR, TNR

S3. Literature examples

	DOI	Location	TP	FN	TN	FP	N	Citations
1	10.1080/10629360903278800	Table 2	5	0	3	0	8	10
2	10.1021/ci200579f	Table 3	10	0	3	1	14	48
3	10.1021/ci020045	Table 5	6	0	7	1	14	51
4a	10.1155/2015/485864	Table 4	5	1	10	1	17	10
4b	10.1155/2015/485864	Table 5	4	2	10	1	17	10
5a	10.1016/j.ejmech.2010.11.029	Table 6	16	1	3	2	22	86
5b	10.1016/j.ejmech.2010.11.029	Table 10	8	9	4	1	22	86
6a	10.1016/j.vascn.2014.07.002	Table 2	2	12	19	1	34	77
6b	10.1016/j.vascn.2014.07.002	Table 3	10	4	20	0	34	77
7a	10.5935/0103-5053.20130066	Table 2	26	0	6	2	34	61
7b	10.5935/0103-5053.20130066	Table 3	24	2	6	2	34	61
8	10.1016/j.scitotenv.2018.05.081	Table 2	28	9	3	4	44	18
9a	10.4314/wsa.v36i4.58411	Table 2	19	3	18	10	50	14
9b	10.4314/wsa.v36i4.58411	Table 2	21	1	20	8	50	14
10	10.1016/j.bspc.2017.01.012	Figure 2	31	5	24	4	64	80
11	10.1039/C7MD00633K	Figure 3	40	7	15	8	70	9
12	10.3389/fnins.2018.01008	Figure 3	31	9	20	13	73	1
13a	10.4315/0362-028X-61.2.221	Table 3	79	14	19	0	112	52
13b	10.4315/0362-028X-61.2.221	Table 3	89	4	16	3	112	52
14a	10.1016/j.ancr.2014.06.005	Figure 6.3	136	2	2	12	152	7
15a	10.1016/j.saa.2016.09.028	Table 2	3	12	150	0	165	65
15b	10.1016/j.saa.2016.09.028	Table 2	6	9	150	0	165	65
16	10.1021/acs.analchem.7b00426	Table 3	188	0	13	2	203	28
14b	10.1016/j.ancr.2014.06.005	Table 3	253	27	11	59	350	7

Table S1. Literature examples of classifiers with small sample size (N). Citations were recorded on Google Scholar on June 16th, 2020 at 12:55 pm.

S4. Proof that variance of metric distributions calculated from synthetic confusion matrices is systematically too large

For a confusion probability matrix (θ) following a Dirichlet distribution with parameter vector α

$$\theta \sim \text{Dirichlet}(\alpha) \tag{S1}$$

where α is the sum of the CM and the prior, the expected value and variance are

$$\text{E}[\theta_i] = \frac{\alpha_i}{\alpha_0} \tag{S2}$$

$$\text{Var}[\theta_i] = \frac{\alpha_i}{\alpha_0}\left(\frac{1-\frac{\alpha_i}{\alpha_0}}{1+\alpha_0}\right) \tag{S3}$$

where $\alpha_0 = \sum \alpha_k$. The expected value and variance of the entry V_i of a confusion matrix generated by a multinomial distribution

$$V = [V_{\text{TP}}, V_{\text{FN}}, V_{\text{TN}}, V_{\text{FP}}] \sim \text{Multinomial}(\theta, N) \tag{S4}$$

is given by

$$\text{E}[V_i] = N\frac{\alpha_i}{\alpha_0} = N\text{E}[\theta_i] \tag{S5}$$

$$\text{Var}[V_i] = N(N+\alpha_0)\frac{\alpha_i}{\alpha_0}\left(\frac{1-\frac{\alpha_i}{\alpha_0}}{1+\alpha_0}\right) = N(N+\alpha_0)\text{Var}[\theta_i] \tag{S6}$$

From this, we can calculate the expected value and variance for the proportion of i, $\frac{V_i}{N}$

$$\text{E}\left[\frac{V_i}{N}\right] = \frac{1}{N}\text{E}[V_i] = \text{E}[\theta_i] \tag{S7}$$

$$\text{Var}\left[\frac{V_i}{N}\right] = \frac{1}{N^2}\text{Var}[V_i] = \left(1+\frac{\alpha_0}{N}\right)\text{Var}[\theta_i] \tag{S8}$$

Whereas $\text{E}\left[\frac{V_i}{N}\right]$ is independent of N, $\text{Var}\left[\frac{V_i}{N}\right]$ is not. In Caelen's approach, $N \approx \alpha_0$. Therefore, the variance will be overestimated by approximately a factor of two. Since the variance of $\frac{V_i}{N}$ are overestimated w.r.t. θ_i, the same holds for $\frac{V}{N}$ w.r.t. θ and metrics calculated on $\frac{V}{N}$ and θ, respectively.

If N was increased beyond α_0, it would converge towards the true variance

$$\lim_{N\to\infty} \text{Var}\left[\frac{V_i}{N}\right] = \text{Var}[\theta_i]\,. \tag{S9}$$

S5. Rule for sample size determination of metrics modeled by a beta distribution

For a normal distribution, approximately 95% of the density are within two standard deviations σ from the mean. Therefore, the length of the 95% highest posterior density interval will be close to 4σ. According to the central limit theorem, beta distributions behave for large sample sizes like normal distributions. The standard deviation σ of a beta distribution is given by

$$\sigma = \sqrt{\frac{\alpha\cdot\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}}. \tag{S10}$$

where α and β are the counts of observations per class, where the meaning of "class" depends on the studied metric. As discussed in the main text, if one is looking at accuracy (ACC), α denotes correct classifications (TP + TN) and β denotes wrong classifications (FP + FN). In the case of TPR, α counts the number of true positives (TPs) whereas β counts false negatives (FNs).

To make explicit the dependency on sample size N, we express α as $a\cdot N$ and β as $b\cdot N$ with fractions $a = \frac{\alpha}{N}, b = \frac{\beta}{N}$ of the two classes.

$$\sigma = \sqrt{\frac{a\cdot N\cdot b\cdot N}{(a\cdot N+b\cdot N)^2(a\cdot N+b\cdot N+1)}} \tag{S11}$$

$$\sigma = \sqrt{\frac{N^2\cdot a\cdot b}{N^2(a+b)^2(N(a+b)+1)}} \tag{S12}$$

$$\sigma = \sqrt{\frac{a\cdot b}{(a+b)^2(N(a+b)+1)}} \tag{S13}$$

Since $\alpha + \beta = N$, we know that $a + b = 1$. Now we can simplify Eq (S13) to

$$\sigma = \sqrt{\frac{a \cdot b}{N + 1}} \quad [S14]$$

For large N, this approximates to

$$\sigma \approx \sqrt{\frac{a \cdot b}{N}} \quad [S15]$$

σ is largest if $a = b = 0.5$.

$$\sigma_{max} \approx \sqrt{\frac{0.5 \cdot 0.5}{N}} \quad [S16]$$

$$\sigma_{max} \approx \frac{0.5}{\sqrt{N}} \quad [S17]$$

In the main text, we have defined metric uncertainty (MU) as the length of the 95% highest posterior density interval. Therefore, its upper limit can be approximated as $4\sigma \approx \frac{2}{\sqrt{N}}$. If one cannot reject the possibility that $a = b = 0.5$, one will need $\frac{4}{MU^2}$ samples to obtain the desired MU.

References

1. Fisher R (1922) On the interpretation of $\chi 2$ from contingency tables, and the calculation of P. *J. R. Stat. Soc.* 85(1):87–94.
2. McElreath R (2018) *Statistical rethinking: A Bayesian course with examples in R and Stan.* (Chapman and Hall/CRC).
3. Gelman A (2003) A Bayesian Formulation of Exploratory Data Analysis and Goodness-of-fit Testing*. *Int. Stat. Rev.* 71(2):369–382.

Chapter 4

Conclusions & Outlook

4.1 Understanding Autolysis Acceleration

Accelerated self-digest, first observed almost 50 years ago, [7] has been rediscovered recently and was proposed as a protease inhibition strategy with therapeutic potential. [10] Previously known accelerators, such as dextrane sulfate, rely on unspecific electrostatic interactions and lose their inhibitory power at physiological salt concentrations. The affinity copolymers presented in section 3.1 bind to amino acids on trypsin's surface and drastically reduce its proteolytic activity under physiological conditions. Therefore, they are more promising candidates for the development of therapeutic inhibitors to tackle pathological misregulation of proteases, as is the case in patients suffering from pancreatitis. Another advantage of the presented copolymers is their composition out of amino acid-selective monomers. Based on a combinatorial library of such monomers, copolymers can be tailored to protease surfaces with relative ease in the future. [17] Multiple analytical methods proved that diminished trypsin activity is the result of autolysis. Computational work indicated that the polymer does not cover the active site nor all of the cleavage sites on trypsin's surface. These two findings are essential to the hypothesis that the protease concentration is elevated on the polymer's surface, causing the acceleration in autolysis.

This theory of local enrichment contrasts with Johnson and Whateley's hypothesis of increased proteolytic susceptibility of trypsin upon adsorption on colloidal silica. [8] In their paper, Johnson and Whateley argue that their hypothesis fits the first-order reaction they observe. Their reasoning relies on the assumptions that 1) the colloid surface is fully covered with trypsins and 2) autolysis in solution is much slower than on the surface and can be neglected. Surprisingly, they did not verify if these assumptions are in agreement with the adsorption isotherme they presented or the kinetic data on regular trypsin self-digest. We developed a Bayesian approach to integrate the adsorption data with the kinetic data on trypsin autolysis with and without the silica present in a single mathematical model. Our findings demonstrate that Johnson and Whateley's original hypothesis can not be correct because the silica surface is not fully covered with trypsins and because the autolysis in solution is non-negligible. Most likely, trypsin is present in two conformations A and B, both of which must be modeled, where conformation A digests conformation B. Refining the kinetic models with this feature, both hypotheses, assuming increased susceptibility upon adsorption or local enrichment, can explain the data well. The conversion from conformation A to conformation B, which is not impacted by the presence of silica, is the rate-determining step resulting in first-order kinetics. Therefore, the mechanistic details of the surface reaction cannot be identified. Neither theory can be rejected based on the information that is currently available.

Occam's razor, stating that the simplest explanation is most likely the right one,

would favor the hypothesis that local enrichment is responsible for accelerated trypsin degradation. It has been proven experimentally that polymers and silica surfaces can bind multiple trypsins simultaneously. The only additional assumption necessary is that those trypsins remain active and mobile upon adsorption. Johnson and Whateley found that adsorbed trypsin is equally active; our computational work in section 3.1 showed that the active site is likely to remain accessible if the polymer binds to trypsin's surface. Moreover, trypsins bound to polymers can be assumed to remain mobile due to the polymers' flexibility. Computational work has indicated the formation of protein rich phases around polymers. [39] In this case, local concentration would be high although only a part of the trypsins is bound to the polymer. The other proteases would be associated to the bound proteases and even more mobile. This could also be the case for colloids. Additionally, it is credible that trypsins can (partially) desorb from silica surfaces to digest neighboring proteins.

Johnson's hypothesis on the other hand depends completely on the idea that adsorption increases susceptibility. While this assumption is not implausible, and has been proposed similarly for savinase, [9, 40] another serine protease, I am not aware of evidence that this is the case for trypsin. Although conformational changes upon adsorption have been observed, it is unclear if they can be linked to increased susceptibility (section 4.1.1). Therefore, Johnson's hypothesis requires a somewhat stronger assumption than the one presented in section 3.1. It is also conceivable that colloidal silica and affinity copolymers accelerate autolysis in different ways or that both pathways contribute.

Although we could not identify the reaction mechanism of surface accelerated trypsin autolysis definitely through kinetic modeling due to insufficient available data, it is a powerful tool and helped to improve our mechanistic understanding. It links a microscopically detailed reaction hypothesis to macroscopic data. Reactions and most assumptions must be modeled explicitly, which is a great benefit. Moreover, the Bayesian data integration approach we developed is a valuable result in its own right. It can easily be extended to new data or different studies. While combinations of kinetic models with thermodynamic information are the most straightforward, virtually any data set can be incorporated. The only requirement is that the probabilistic models, describing the individual data sets, share at least one parameter.

4.1.1 Potential Experiments to Identify the Reaction Mechanism

Definitive identification of the mechanism behind accelerated self-digest would be highly desirable not only for protease regulation, but also because the approach could be generalized towards any digester-substrate pair. This way, all enzymatic reactions could be accelerated. The potential therapeutic and biotechnological applications of such accelerators could be vast.

Use of Atomistic Modeling is Limited

Our kinetic study suggests that at least two trypsin conformations are involved in autolysis. One of them, called conformation B, is presumed to be more susceptible to cleavage. Molecular modeling of surface or polymer accelerated self-digest would require a decent structure of conformation B because of its critical role. It is unlikely that such a structure could be obtained either experimentally or computationally. The vast majority of trypsin structures found on the protein data base (PDB) contain Ca^{2+}. Ca^{2+} is known to prevent autolysis, presumably due to a shift of the equilibrium away from conformation B towards conformation A, which is more resistant to autolysis. [41] Figure 6b in section 3.2 shows that even without Ca^{2+}, conformation A is preferred. I compared the PDB structures

without Ca^{2+} to those with the ion binding and found no obvious structural differences close to well known cleavage sites such as R117 (results not shown). It is plausible that the more populated and stable conformation A crystallizes whereas conformation B does not. During crystallization, the local concentration of trypsins is high. Unless autolysis is prevented, conformation B would likely be cleaved rapidly under such conditions.

In principle, conformation B can be sampled using Molecular Dynamics (MD) simulations. A structure from the PDB, presumably representative of conformation A, would serve as a starting point. Since classic MD simulations do not allow to observe breakage of the peptide bond (this would require quantum mechanics or a reactive force field, further increasing the computational cost), it is unclear how one could identify a more susceptible conformation. Two conformational features have been linked to susceptibility to proteolysis, namely high accessible surface area [42] and flexibility [43] of relevant amino acids. Both increase the likelihood that the cleavage site of the substrate can be in vicinity of the digester's active site. Still, these criteria are rather vague and there is no guarantee that the digester would not be repelled by the surrounding of such an exposed, flexible residue.

In 2005, Papaleo et al. ran multiple 6 ns MD simulations of trypsin with Ca^{2+} and of a trypsin structure where the Ca^{2+} had been removed. [44] Potentially, the structure without Ca^{2+} is more likely to represent conformation B. Some of their findings are counterintuitive, for example R117 is less accessible if Ca^{2+} is absent, yet it has been shown to be a critical cleavage site. [45, 46] The most likely explanation is that their trajectories were too short. Nanoseconds are not enough to reliably sample conformational differences for proteins. Observed differences between the conformations with and without Ca^{2+} might be statistical flukes. Due to the tremendous improvements both in soft- and hardware, trajectories can be orders of magnitudes longer today. We ran MD simulations spanning several microseconds in total of trypsin with and without Ca^{2+}. We found that R117 is substantially more exposed if there is no Ca^{2+} (unpublished). Yet, it is unclear if this by itself increases proteolytic susceptibility or if this is only the case if a larger rearrangement occurs.

Furthermore, our trajectories are much too short to observe conformation B. Figure 6b in section 3.2 shows that the conversion from A to B is the slowest step. Its rate k_{AB} is of the order of magnitude of 10^{-4} s^{-1}. The mean lifetime of conformation A is therefore $\frac{1}{k_{\mathrm{AB}}} \approx 10^4$ s ≈ 3 h. The conformational change is therefore expected to occur spontaneously on a time span that is approximately one billion times longer than our MD simulations. It seems implausible that conformation B can be observed computationally in the next years, even if highly specialized hardware and/or advanced sampling techniques are available. [47] Atomistic molecular modeling is not a promising approach to determine which of the hypotheses discussed in section 3.2 is more likely. Consequently, bottom-up approaches to build coarser models are also severely limited.

Probe Structural Changes Upon Adsorption

Another idea would be to study if trypsin changes its conformation upon adsorption on the silica surface. If this is the case, we know from the kinetic model that this step must be faster than the rate determining conversion of A to B in the solvent. The adsorption induced conformation change might still be orders of magnitude slower than what can be observed using MD simulations.

Experiments monitoring conformational changes upon adsorption can cover longer time scales than atomistic simulations. Koutsopoulos et al. compared the conformation of trypsin in solution to trypsin adsorbed on silica wafers. [48] They found structural

differences with circular dichroism and fluorescence spectroscopy, yet, the results are not detailed enough to evaluate if these changes are linked to proteolytic susceptibility. In order to verify Johnson and Whateley's hypothesis of increased susceptibility upon adsorption, more conformational details would be needed. It is unclear which experiments could provide them.

Load Silica Surface with Deactivated Trypsins and Expose It to Active Protease

Ideally, there would be an experiment that tests the mechanistic hypotheses directly rather than their underlying assumptions. A kinetic experiment could look like this:

1. Load silica with an inactivated trypsin mutant, for example S195A. [49]
2. Remove supernatant. If the silica is colloidal, centrifugation may separate the loaded colloids from most of the supernatant. If silica wafers were used, as in reference [50], these can be removed from the supernatant. The surface load needs to be verified after this step.
3. Introduce loaded silica into solution containing active trypsin.
4. Take samples from the solution and test them for autolysis fragments or monitor the surface coverage of silica.

A similar experiment could be designed for polymers. The separation of the loaded polymers from the supernatant in step 2 might be more challenging, though. If Johnson and Whateley's hypothesis of increased susceptibility of adsorbed trypsin is correct, degradation of adsorbed trypsins should start immediately. On the other hand, if increased local concentration causes the acceleration in self-digest, autolysis on the surface would be delayed. Since the surface is fully covered with inactive trypsin molecules, these would need to desorb first to free up space for active trypsins to bind before the surface reaction could begin.

The inferred parameters from section 3.2 can be used to model this experiment beforehand. Assuming that 0.242 m^2 of silica (equivalent to 1 mg colloidal Syton X used by Johnson and Whateley) fully loaded with inactivated trypsin is introduced into 1 L of a solution containing 0.5 µM active trypsin, monitoring the surface coverage on short time scales would reveal the mechanism of autolysis acceleration (Figure 4.1). If the Eley-Rideal (ER) like mechanism, which represents Johnson and Whateley's hypothesis of increased proteolytic susceptibility of adsorbed trypsins, is correct, the surface coverage should decrease noticeably in a few seconds or less. Following the Langmuir-Hinshelwood (LH) mechanism, corresponding to the hypothesis of local enrichment, this would take almost an order of magnitude longer. The time resolution would need to be 3 seconds or less. After approximately 10 seconds, the reaction is predicted to continue equally fast following either the Eley-Rideal or Langmuir-Hinshelwood mechanism.

This experiment can only be executed successfully to distinguish between the two hypotheses if the following assumptions and conditions are fulfilled:

1. Inactivated trypsin is comparable in susceptibility and adsorption behavior to regular trypsin.
2. Adsorbed trypsins are not impacted by the removal of the supernatant (step 2 in the description of the experiment).

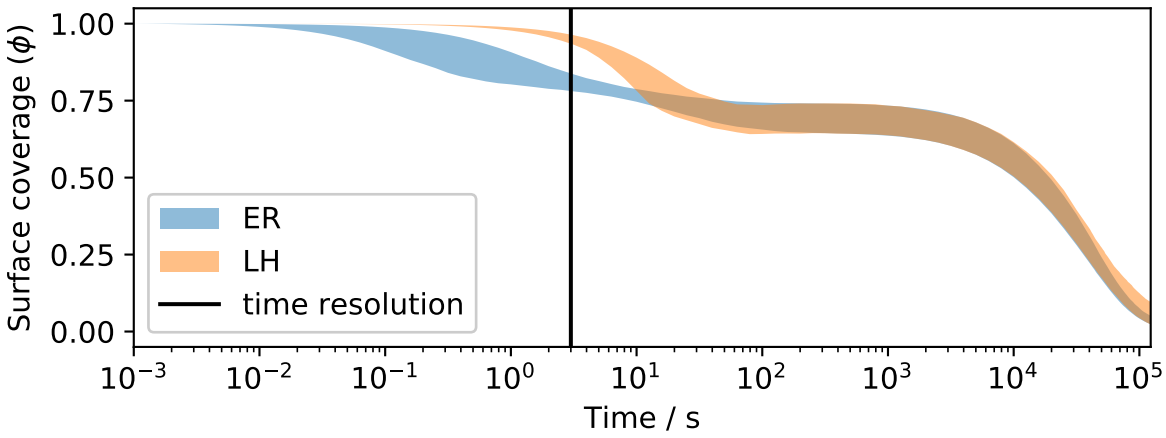


Figure 4.1: Surface coverage ϕ of a silica surface that is loaded with inactivated trypsin but later transferred into a solution of active trypsins (see text) over time. The coverage is modeled with both the Eley-Rideal (ER) and the Langmuir-Hinshelwood (LH) like models. Reaction constants are identical to those inferred in section 3.2. Areas depict the 16-84% credible intervals. Measurement errors are neglected. The time resolution of ellipsometry, three seconds, is shown as black line.

3. Conformations A and B are in equilibrium in both the regular and inactivated trypsin when the loaded silica is introduced into the solution.

4. The surface coverage can be monitored with a time resolution of a few seconds or even less. This would not be possible with the same experimental setup Johnson and Whateley used for their study. Ellipsometry has been employed to study trypsin adsorption on silica wafers, the time resolution is said to be approximately 3 seconds. [50] This might be just fast enough to observe the difference.

5. Introducing the loaded colloids into the trypsin solution, inhomogeneity effects can be neglected.

6. The materials and experimental conditions are identical to Johnson and Whateley's study to ensure transferability of the inferred parameters. It is not guaranteed that reaction constants for adsorption, desorption, and the surface reaction are appropriate if silica wafers are used instead of colloidal silica.

4.1.2 Surfaces and Polymers to Decrease Autolysis

Interestingly, surfaces and polymers have also been suggested to protect proteases from autolysis. Koutsopoulos et al. state that trypsins adsorbed on silica wafers are 90% less active catalytically which would reduce the rate of self-digest. [48] They assumed that conformational changes are responsible for this drop (see section 4.1.1). In contrast, Johnson and Whateley's paper explicitly states that adsorbed trypsin is equally active compared to its unbound form. While it could be that silica wafers reduce trypsin's activity whereas colloidal silica does not, Hahn et al. observe depletion of silica wafers covered with trypsins over time, presumably due to autolysis. [51] Perhaps it is more plausible that Koutsopoulos' results are in fact caused by accelerated self-digest rather than a structural rearrangement that reduces the enzymatic activity directly. In their experiments, they placed silica wafers in a trypsin solution and compared its catalytic

activity to a control solution without silica. If these experiments were not executed quickly, autolysis decreased the number of intact trypsins faster with silica present, leading to a lower observed activity.

Due to the suspected reduction in catalytic activity, Koutsopoulos et al. propose to store trypsin adsorbed on insoluble particles such as silica wafers in order to protect it from autolysis. This strategy is the opposite of exposing trypsin to silica surfaces to accelerate its autolysis. Therefore, it does not seem promising without further proof that the reduction in activity observed by Koutsopoulos is indeed caused by structural changes rather than autolysis.

In another study, Sasai demonstrated polymers that can effectively protect trypsin from self-digest, allowing to store it for months. [52] These polymers bind covalently to the protease surface. Sasai et al. assumed that the protease surfaces covered by the negatively charged polymer repel each other electrostatically. In contrast, we assume in section 3.1 that negatively charged polymers act as a template for trypsins, enriching them locally. If Sasai's polymers are present in solution but do not bind to trypsins covalently, autolysis is much faster than if the polymer is absent. Apparently, the type of bond between polymer and trypsin has a dramatic effect on autolysis, shifting self-digest from one extreme (acceleration if the bond is non-covalent) to the other (reduction if the bond is covalent). The authors did not discuss this discrepancy in their publication. Their polymer contains N-hydroxysuccinimide esters. These are known to react with the primary amine on lysine's side chain. [53] Lysines are important cleavage sites for trypsin; in fact, trypsin's lysine residues are commonly methylated to prevent autolysis. [54] Therefore, it seems plausible that the reduction in autolysis observed by Sasai et al. is caused by the protection of the lysines rather than repulsion between trypsins.

4.2 Classifier Evaluation

In addition to the work on accelerated protease autolysis, this thesis contains contributions towards the understanding of classifier performance.

4.2.1 Metric Uncertainty Is Larger Than Regularly Assumed

We demonstrated that metric uncertainty (MU) is large for sample sizes below 1000, which are common in biology and chemistry (section 3.3). The uncertainty of key metrics for published classifiers sometimes exceeds 60 percentage points. Some classifiers might even be worse than random guessing. In order to enable both classifier developers and potential users to determine the uncertainty, we published our browser-based tool. Moreover, we derived a rule that allows to determine the sample size beforehand, which is helpful during the design of classification studies.

Our work shows how difficult it is to identify the best classifier out of a pool of options. Uncertainty in the metrics due to limited sample sizes can blur the distinctions if the sample size and the performance differences between the classifiers are small. Yet, one should not despair if there is no classifier that is obviously the best. In many instances, there will be one or more classifiers that are good enough. These would be decent choices, even if one cannot decisively say which one is the best.

Increase Representativity of Metric Uncertainty Paper Using Meta-Study

One of the shortcomings of section 3.3 is that the literature examples of classifiers with small test sets do not allow for a representative assessment of the impact of MU in any

given field. Meta-studies allow for a comprehensive overview of the state of the art in a field. My Bayesian approach would allow a probabilistic understanding of how much the classifiers differ, similar to Figure 3 in section 3.3.

Extend Sample Size Determination

The rule for sample size determination in section 3.3 only applies to metrics that can be described with a beta distribution. Therefore, it is not applicable to more complex metrics such as Matthews correlation coefficient or Bookmaker informedness. One could generate empirical estimates of metric uncertainty for varying sample sizes for these metrics, which would give users useful orientation for the design of their classification experiments.

4.2.2 Classification Is an Unnecessary Binarization

Although section 3.3 is helpful in classification tasks, one should avoid classification wherever possible for two reasons.

1) In many cases, the classes do not exist. They are a simplification by the modeler rather than a physical fact. Regression would be more appropriate. For example, protease inhibition is obviously non-binary: we know well that different substances inhibit enzymes with varying strength (Figure 4.2). Assigning these substances to the classes *inhibitor* and *non-inhibitor* discards information about the differences between them. Moreover, it will in many cases be subjective. Compound C in Figure 4.2 decreases protease activity but only at relatively high concentrations. Moreover, it can reduce the activity to 60% at most. Consequently, compound C is perhaps not promising if one is looking for a potent inhibitor yet its interference with the protease is also non-negligible. Considering compound C an inhibitor or non-inhibitor is therefore subjective.

In some cases however, classification models might be the only option since regression models cannot be build. For example, publications often only report experimental half maximal inhibitory concentration (IC_{50}) values. For non-inhibitors or weak inhibitors, IC_{50} values cannot be determined (green and red lines in Figure 4.2). The only adequate description of this would be to assume that IC_{50} equals infinity because an infinite amount of the substance is needed to decrease the protease activity by 50%. This cannot be modeled with regression, therefore classification would be the only option if IC_{50} values are the only available data.

2) Even if classes are reasonable, probabilities matter. Being dead or alive is inherently binary in most scenarios. Nevertheless, one would prefer to know the probability that an individual is dead rather than just the binary label *dead*. Imagine a rescue crew in a mountain village after an avalanche. The team relies on an algorithm to predict which inhabitants are still alive and therefore require immediate help. Giving them a list with people who are predicted to be alive is helpful. It would be much better though to give them a list with scores (or better yet: well calibrated probabilities) indicating which inhabitants are more likely than others to be alive. Such a list would allow for a much more informed prioritization of help in the given scenario. This example disproves the argument that binary labels are sufficient because decisions would be binary anyway.

Most classification algorithms calculate scores for every class. Application of a threshold binarizes these scores into labels. This step discards valuable information, decreasing the knowledge gained from the model.

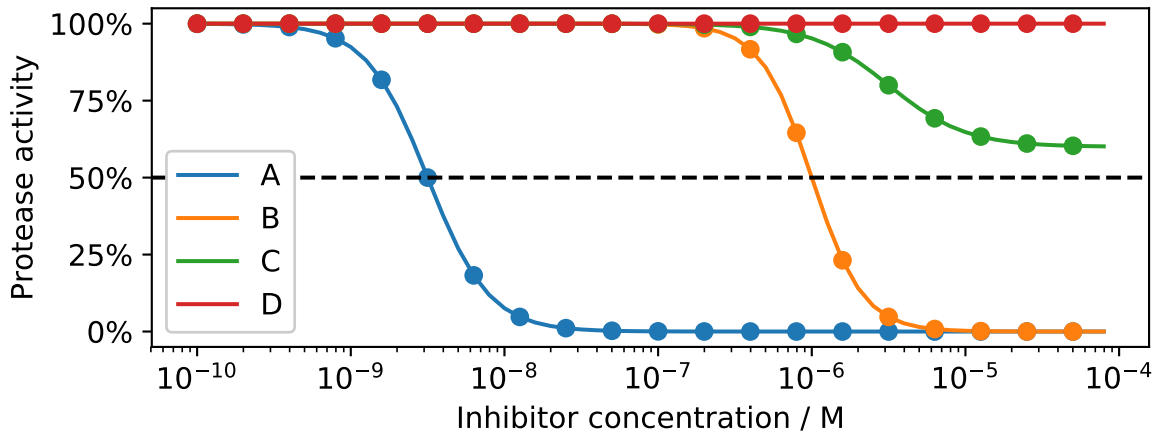


Figure 4.2: Reduction in protease activity is dependent on the concentration of the inhibitor. Since different molecules inhibit proteases with varying strength, protease inhibition is a regression problem rather than classification. The half maximal inhibitory concentration is defined as the concentration of inhibitor where 50% of the original activity remains (dashed line).

Class labels are used for two main reasons even if scores are available. Labels are somewhat easier to understand and perhaps more intuitive than scores. Moreover, convention is an important factor. A lot of teaching material instructs users to apply thresholds to their scores to binarize results to class labels. Classification is so established that one often does not consider the associated approximations and simplifications.

Summarizing, class labels should be avoided if scores are available or, which would be even better, regression is possible. While this might seem to decrease the importance of section 3.3, lots of appropriate and inappropriate discretizations will continue to be published. Our article might contribute to a more realistic understanding of classifier performance, independent of the question if classification was the appropriate choice to begin with.

Bibliography

[1] Rogers, L. D & Overall, C. M. (2013) Proteolytic post-translational modification of proteins: Proteomic tools and methodology. *Mol. Cell. Proteomics* **12**, 3532–3542.

[2] Hirota, M, Ohmuraya, M, & Baba, H. (2006) The role of trypsin, trypsin inhibitor, and trypsin receptor in the onset and aggravation of pancreatitis. *J. Gastroenterol.* **41**, 832–836.

[3] Waller, A, Long, B, Koyfman, A, & Gottlieb, M. (2018) Acute Pancreatitis: Updates for Emergency Clinicians. *J. Emerg. Med.* **55**, 769–779.

[4] Kirk, O, Borchert, T. V, & Fuglsang, C. C. (2002) Industrial enzyme applications. *Curr. Opin. Biotechnol.* **13**, 345–351.

[5] Belorgey, D, Dirrig, S, Amouric, M, Figarella, C, & Bieth, J. G. (1996) Inhibition of human pancreatic proteinases by mucus proteinase inhibitor, eglin c and aprotinin. *Biochem. J.* **313**, 555–560.

[6] Kunitz, M & Northrop, J. H. (1934) Inactivation of Crystalline Trypsin. *J. Gen. Physiol.* **17**, 591–615.

[7] Johnson, P & Whateley, T. (1972) The effect of glass and silica surfaces on trypsin and α-chymotrypsin kinetics. *Biochim. Biophys. Acta - Enzymol.* **276**, 323–327.

[8] Johnson, P & Whateley, T. L. (1981) The effects of surface and macromolecular interactions on the kinetics of inactivation of trypsin and alpha-chymotrypsin. *Biochem. J.* **193**, 285–94.

[9] Maste, M. C, Rinia, H. A, Brands, C. M, Egmond, M. R, & Norde, W. (1995) Inactivation of a subtilisin in colloidal systems. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **1252**, 261–268.

[10] Lv, Y, Zhang, J, Wu, H, Zhao, S, Song, Y, Wang, S, Wang, B, Lv, G, & Ma, X. (2015) A protease inhibition strategy based on acceleration of autolysis. *Chem. Commun.* **51**, 5959–5962.

[11] Fasting, C, Schalley, C. A, Weber, M, Seitz, O, Hecht, S, Koksch, B, Dernedde, J, Graf, C, Knapp, E. W, & Haag, R. (2012) Multivalency as a chemical organization and action principle. *Angew. Chemie - Int. Ed.* **51**, 10472–10498.

[12] Schrödinger, L. (2019) The PyMOL Molecular Graphics System, Version 2.3.4.

[13] Scheidig, A. J, Hynes, T. R, Pelletier, L. A, Wells, J. A, & Kossiakoff, A. A. (1997) Crystal structures of bovine chymotrypsin and trypsin complexed to the inhibitor domain of Alzheimer's amyloid β-protein precursor (APPI) and basic pancreatic trypsin inhibitor (BPTI): Engineering of inhibitors with altered specificities. *Protein Sci.* **6**, 1806–1824.

[14] Hanson, W. M, Domek, G. J, Horvath, M. P, & Goldenberg, D. P. (2007) Rigidification of a flexible protease inhibitor variant upon binding to trypsin. *J. Mol. Biol.* **366**, 230–243.

[15] Søndergaard, C. R, Olsson, M. H. M, Rostkowski, M, & Jensen, J. H. (2011) Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of p K a Values. *J. Chem. Theory Comput.* **7**, 2284–2295.

[16] Olsson, M. H. M, Søndergaard, C. R, Rostkowski, M, & Jensen, J. H. (2011) PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical p K a Predictions. *J. Chem. Theory Comput.* **7**, 525–537.

[17] Gilles, P, Wenck, K, Stratmann, I, Kirsch, M, Smolin, D. A, Schaller, T, De Groot, H, Kraft, A, & Schrader, T. (2017) High-Affinity Copolymers Inhibit Digestive Enzymes by Surface Recognition. *Biomacromolecules* **18**, 1772–1784.

[18] Ohl, L. (2018) Ph.D. thesis (University of Duisburg-Essen).

[19] Pierce, L. J, Strickland, D. J, & Smith, E. S. (1990) The case of Ohio v. Robinson. An 1870 bite mark case. *Am. J. Forensic Med. Pathol.* **11**, 171–177.

[20] Committee on Identifying the Needs of the Forensic Sciences Community, N. R. C. (1986) *Strengthening Forensic Science in the United States: A Path Forward.* (National Academies Press) Vol. 31, pp. 69–70.

[21] Luque, A, Carrasco, A, Martín, A, & de las Heras, A. (2019) The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognit.* **91**, 216–231.

[22] Kruschke, J. K. (2015) in *Doing Bayesian Data Anal.* (Elsevier), pp. 99–120.

[23] Brodersen, K. H, Ong, C. S, Stephan, K. E, & Buhmann, J. M. (2010) The balanced accuracy and its posterior distribution. *Proc. - Int. Conf. Pattern Recognit.* pp. 3121–3124.

[24] Caelen, O. (2017) A Bayesian interpretation of the confusion matrix. *Ann. Math. Artif. Intell.* **81**, 429–450.

[25] Fisher, R. (1922) On the interpretation of $\chi 2$ from contingency tables, and the calculation of P. *J. R. Stat. Soc.* **85**, 87–94.

[26] Dietterich, T. G. (1998) Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. *Neural Comput.* **10**, 1895–1923.

[27] Zimmerman, S. B & Trach, S. O. (1991) Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli. *J. Mol. Biol.* **222**, 599–620.

[28] Carpenter, B, Gelman, A, Hoffman, M. D, Lee, D, Goodrich, B, Betancourt, M, Brubaker, M. A, Guo, J, Li, P, & Riddell, A. (2017) Stan: A probabilistic programming language. *J. Stat. Softw.* **76**.

[29] Salvatier, J, Wiecki, T. V, & Fonnesbeck, C. (2016) Probabilistic programming in Python using PyMC3. *PeerJ Comput. Sci.* **2**, e55.

[30] Goodman, J & Weare, J. (2010) Ensemble samplers with affine invariance. *Commun. Appl. Math. Comput. Sci.* **5**, 65–80.

[31] Gelman, A & Loken, E. (2014) The Statistical Crisis in Science. *Am. Sci.* **102**, 460.

[32] Wasserstein, R. L & Lazar, N. A. (2016) The ASA Statement on p -Values: Context, Process, and Purpose. *Am. Stat.* **70**, 129–133.

[33] Ioannidis, J. P. A. (2005) Why Most Published Research Findings Are False. *PLoS Med.* **2**, e124.

[34] Baker, M. (2016) Is there a reproducibility crisis? *Nature* **533**, 353–366.

[35] Wasserstein, R. L, Schirm, A. L, & Lazar, N. A. (2019) Moving to a World Beyond $p < 0.05$. *Am. Stat.* **73**, 1–19.

[36] Smolin, D, Tötsch, N, Grad, J.-N, Linders, J, Kaschani, F, Kaiser, M, Kirsch, M, Hoffmann, D, & Schrader, T. (2020) Accelerated trypsin autolysis by affinity polymer templates. *RSC Adv.* **10**, 28711–28719.

[37] Tötsch, N & Hoffmann, D. (2020) Bayesian Data Integration Questions Classic Study on Protease Self-Digest Kinetics. *ACS Omega* **5**, 15162–15168.

[38] Tötsch, N & Hoffmann, D. (2020) Classifier uncertainty: evidence, potential impact, and probabilistic treatment. *arXiv* p. arxiv:2006.11105.

[39] Zumbro, E, Witten, J, & Alexander-Katz, A. (2019) Computational Insights into Avidity of Polymeric Multivalent Binders. *Biophys. J.* **117**, 892–902.

[40] Maste, M. C, Norde, W, & Visser, A. J. (1997) Adsorption-induced conformational changes in the serine proteinase savinase: A tryptophan fluorescence and circular dichroism study. *J. Colloid Interface Sci.* **196**, 224–230.

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

[42] Novotný, J & Bruccoleri, R. E. (1987) Correlation among sites of limited proteolysis, enzyme accessibility and segmental mobility. *FEBS Lett.* **211**, 185–189.

[43] Fontana, A, Fassina, G, Vita, C, Dalzoppo, D, Zamai, M, & Zambonin, M. (1986) Correlation between Sites of Limited Proteolysis and Segmental Mobility in Thermolysin. *Biochemistry* **25**, 1847–1851.

[44] Papaleo, E, Fantucci, P, & De Gioia, L. (2005) Effects of calcium binding on structure and autolysis regulation in trypsins. A molecular dynamics investigation. *J. Chem. Theory Comput.* **1**, 1286–1297.

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

[46] Whitcomb, D. C, Gorry, M. C, Preston, R. A, Furey, W, Sossenheimer, M. J, Ulrich, C. D, Martin, S. P, Gates, L. K, Amann, S. T, Toskes, P. P, Liddle, R, McGrath, K, Uomo, G, Post, J. C, & Ehrlich, G. D. (1996) Hereditary pancreatitis is caused by a mutation in the cationic trypsinogen gene. *Nat. Genet.* **14**, 141–145.

[47] Noé, F. (2015) Beating the millisecond barrier in molecular dynamics simulations. *Biophys. J.* **108**, 228–229.

[48] Koutsopoulos, S, Patzscly, K, Bosker, W. T, & Norde, W. (2007) Adsorption of trypsin on hydrophilic and hydrophobic surfaces. *Langmuir* **23**, 2000–2006.

[49] Zakharova, E, Horvath, M. P, & Goldenberg, D. P. (2009) Structure of a serine protease poised to resynthesize a peptide bond. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 11034–11039.

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

[51] Hahn, T, Marquand, A. F, Plichta, M. M, Ehlis, A. C, Schecklmann, M. W, Dresler, T, Jarczok, T. A, Eirich, E, Leonhard, C, Reif, A, Lesch, K. P, Brammer, M. J, Mourao-Miranda, J, & Fallgatter, A. J. (2013) A novel approach to probabilistic biomarker-based classification using functional near-infrared spectroscopy. *Hum. Brain Mapp.* **34**, 1102–1114.

[52] Sasai, Y, Kanno, H, Doi, N, Yamauchi, Y, Kuzuya, M, & Kondo, S.-i. (2016) Synthesis and Characterization of Highly Stabilized PolymerTrypsin Conjugates with Autolysis Resistance. *Catalysts* **7**, 4.

[53] Hermanson, G. T. (2013) *Bioconjugate techniques.* (Academic press).

[54] Rice, R. H, Means, G. E, & Brown, W. D. (1977) Stabilization of bovine trypsin by reductive methylation. *Biochim. Biophys. Acta (BBA)-Protein Struct.* **492**, 316–321.

List of Figures

List of Tables

Glossary

D data

N sample size

θ parameters

BM Bookmaker informedness

CM confusion matrix

FN false negative

FP false positive

IC_{50} half maximal inhibitory concentration

MCC Matthews correlation coefficient

NHST null-hypothesis significance testing

NPV negative predictive value

PPV positive predictive value

TNR true negative rate

TN true negative

TPR true positive rate

TP true positive

AS accelerated self-digest

ER Eley-Rideal

LH Langmuir-Hinshelwood

MCMC Markov chain Monte Carlo

MD Molecular Dynamics

MU metric uncertainty

PDB protein data base

Acknowledgments

Writing the last lines of this thesis I look back at the amazing ride the past three years have been. I appreciate that I have been extremely fortunate to work in an environment of smart, motivated, and curious people that on top of that are fun to spend time with. First, I would like to sincerely thank Daniel Hoffmann for his supervision. You always fostered independence and the pursuit of novel ideas. Critically, you helped me convert sometimes vague ideas into hypotheses and ultimately computational experiments. Your emphasis on long-term learning over short-term success reassured me to follow my intuitions and develop my own projects because I knew that even if they led nowhere, the lessons along the way would make up for it. Also, thanks go to Thomas Schrader for co-supervising my PhD and answering my questions about the experimental work my computational studies are based on. I am grateful to Katja Ickstadt who kindly accepted to serve as third reviewer of my thesis.

Thanks a lot to all members of the "Bioinformatics and Computational Physics" group, past and present! I had a great time and I will always remember your willingness to help, our discussions about statistics, politics, movies, weird science and so much more fondly. Ludwig Ohl's support to familiarize myself with the topic as well as the tips and tricks he shared with me about programming, version control, etc. gave me a huge head start. Moreover, thank you for the great time we had sharing an office. Anja Lange made sure I could worry about the science and my own code because the infrastructure had been taken care of. Daniel Habermann, Simo Kitanovski, Martin Theissen, and Christoph Waterkamp made me curious about Bayesian statistics and were a great source of knowledge for me when I started using it myself. Thanks to Jean-Noël Grad for sharing his tremendous knowledge about anything related to molecular modeling and the corresponding literature freely. I enjoyed the long discussions about physical models with Neda Rafieiolhosseini, Marius Welzel, and Florian Kubiak. Thanks to Claudia Wilmes for all her organizational help which allowed me to fully focus on my research. Similarly, I am grateful to all the administrative staff at the university and Sonderforschungsbereich 1093.

Great thanks go out to all my collaborators, section 3.1 and other joint projects are much more than just the sum of their parts thanks to the team work. Daniel Smolin patiently answered the many questions I had about polymer accelerated self-digest when I started my PhD. Paul Bürkner gave me the idea of investigating classification challenges in section 3.3 and Kai Horny's thoughtful feedback on an earlier draft of the manuscript improved the quality and its readability considerably.

If I had had to build everything from scratch, this thesis would have been much thinner. I would like to deeply thank the open-source community. Although one rightfully takes great pride in showcasing a project, there is also the uneasiness of letting others see the imperfections in our work and exposing it to criticism. Thank you for overcoming that uneasiness and letting others build on top of your work. Similarly, I am indebted to the countless people who devote their time and energy sharing their knowledge and helping total strangers to solve their technical problems in online forums.

A thank you to my friends who helped to put the small nuisances of scientific work

and life in perspective. Special thanks go to my dear family, especially my parents, who paved the way to my education and so much more. Finally, my greatest thanks go to my wonderful wife Kristina who has shared my excitement when things were going well and cheered me up when I was grumpy because they were not.

Appendix

Contributions to Publications

Kumulative Dissertation/Beteiligung an Veröffentlichungen

Kumulative Dissertation von Herrn Niklas Tötsch

Autorenbeiträge

Titel der Publikation: Accelerated trypsin autolysis by affinity polymer templates

Autoren: D. Smolin, N. Tötsch, J.-N. Grad, J. Linders, F. Kaschani, M. Kaiser, M. Kirsch, D. Hoffmann, T. Schrader

Anteile:

- Konzept: 10%
- Durchführung der Experimente: 0%
- Datenanalyse: 20%
- Artenanalyse: NA
- Statistische Analyse: 50%
- Manuskripterstellung: 10%
- Überarbeitung des Manuskripts: 15%

Unterschrift Doktorand/in	Unterschrift Betreuer/in

Kumulative Dissertation/Beteiligung an Veröffentlichungen

Kumulative Dissertation von Herrn Niklas Tötsch

Autorenbeiträge

Titel der Publikation: Bayesian Data Integration Questions Classic Study on Protease Self-Digest Kinetics

Autoren: N. Tötsch, D. Hoffmann

Anteile:

- Konzept: 90%
- Durchführung der Experimente: NA
- Datenanalyse: 100%
- Artenanalyse: NA
- Statistische Analyse: 100%
- Manuskripterstellung: 100%
- Überarbeitung des Manuskripts: 60%

Unterschrift Doktorand/in	Unterschrift Betreuer/in

Kumulative Dissertation/Beteiligung an Veröffentlichungen

Kumulative Dissertation von Herrn Niklas Tötsch

Autorenbeiträge

Titel der Publikation: Classifier uncertainty: evidence, potential impact, and probabilistic treatment

Autoren: N. Tötsch, D. Hoffmann

Anteile:

- Konzept: 80%
- Durchführung der Experimente: NA
- Datenanalyse: 100%
- Artenanalyse: NA
- Statistische Analyse: 100%
- Manuskripterstellung: 100%
- Überarbeitung des Manuskripts: 75%

Unterschrift Doktorand/in	Unterschrift Betreuer/in

CV

The curriculum vitae is not included in the online version for data protection reasons.

The curriculum vitae is not included in the online version for data protection reasons.

Declarations

Erklärung:
Hiermit erkläre ich, gem. 7 Abs. (2) d) + f) der Promotionsordnung der Fakultät für Biologie zur Erlangung des Dr. rer. nat., dass ich die vorliegende Dissertation selbständig verfasst und mich keiner anderen als der angegebenen Hilfsmittel bedient, bei der Abfassung der Dissertation nur die angegeben Hilfsmittel benutzt und alle wörtlich oder inhaltlich übernommenen Stellen als solche gekennzeichnet habe.

Essen, den _____ _____
Unterschrift des/r Doktoranden/in

Erklärung:
Hiermit erkläre ich, gem. 7 Abs. (2) e) + g) der Promotionsordnung der Fakultät für Biologie zur Erlangung des Dr. rer. nat., dass ich keine anderen Promotionen bzw. Promotionsversuche in der Vergangenheit durchgeführt habe und dass diese Arbeit von keiner anderen Fakultät/Fachbereich abgelehnt worden ist.

Essen, den _____ _____
Unterschrift des/r Doktoranden/in

Erklärung:
Hiermit erkläre ich, gem. 6 Abs. (2) g) der Promotionsordnung der Fakultät für Biologie zur Erlangung der Dr. rer. nat., dass ich das Arbeitsgebiet, dem das Thema "Insights Into Accelerated Protease Autolysis and Classifier Uncertainty" zuzuordnen ist, in Forschung und Lehre vertrete und den Antrag von Niklas Tötsch befürworte und die Betreuung auch im Falle eines Weggangs, wenn nicht wichtige Gründe dem entgegenstehen, weiterführen werde.

_____
Name des Mitglieds der Universität Duisburg-Essen in Druckbuchstaben

Essen, den _____
Unterschrift eines Mitglieds der Universität Duisburg-Essen