Inferring Proteolytic Processes from Mass Spectrometry Time Series Data Using Degradation Graphs

Stephan Aiche[1,2]*, Knut Reinert[1], Christof Schütte[1], Diana Hildebrand[3], Hartmut Schlüter[3], Tim O. F. Conrad[1]

1 Department of Mathematics and Computer Science, Freie Universität Berlin, Berlin, Germany, 2 International Max Planck Research School for Computational Biology and Scientific Computing, Berlin, Germany, 3 Institute of Clinical Chemistry, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

Abstract

Background: Proteases play an essential part in a variety of biological processes. Besides their importance under healthy conditions they are also known to have a crucial role in complex diseases like cancer. In recent years, it has been shown that not only the fragments produced by proteases but also their dynamics, especially *ex vivo*, can serve as biomarkers. But so far, only a few approaches were taken to explicitly model the dynamics of proteolysis in the context of mass spectrometry.

Results: We introduce a new concept to model proteolytic processes, the *degradation graph*. The degradation graph is an extension of the cleavage graph, a data structure to reconstruct and visualize the proteolytic process. In contrast to previous approaches we extended the model to incorporate endoproteolytic processes and present a method to construct a degradation graph from mass spectrometry time series data. Based on a degradation graph and the intensities extracted from the mass spectra it is possible to estimate reaction rates of the underlying processes. We further suggest a score to rate different degradation graphs in their ability to explain the observed data. This score is used in an iterative heuristic to improve the structure of the initially constructed degradation graph.

Conclusion: We show that the proposed method is able to recover all degraded and generated peptides, the underlying reactions, and the reaction rates of proteolytic processes based on mass spectrometry time series data. We use simulated and real data to demonstrate that a given process can be reconstructed even in the presence of extensive noise, isobaric signals and false identifications. While the model is currently only validated on peptide data it is also applicable to proteins, as long as the necessary time series data can be produced.

Citation: Aiche S, Reinert K, Schütte C, Hildebrand D, Schlüter H, et al. (2012) Inferring Proteolytic Processes from Mass Spectrometry Time Series Data Using Degradation Graphs. PLoS ONE 7(7): e40656. doi:10.1371/journal.pone.0040656

Editor: Emily Parker, University of Cantebury, New Zealand

Received March 8, 2012; **Accepted** June 11, 2012; **Published** July 17, 2012

Funding: The research described in this paper was partially supported by the DFG research center Matheon "Mathematics for key technologies" in Berlin and the BMBF grant 0313694 A. SA was partially supported by the Center for Scientific Simulation, Freie Universität Berlin. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: stephan.aiche@fu-berlin.de

Introduction

Our view of proteases has changed drastically over the last two decades. Once thought to be only associated with simple protein degradation processes they are now recognized to play an important role in a variety of fundamental biological processes across species [1–3]. Furthermore, also their general importance in complex diseases such as cancer or HIV was described [4–8] and they were recognized as possible drug-targets [7,9]. In the last decade, researchers also began to look more closely into the dynamics of proteolytic processes. It was found that changing dynamics of specific protease activity can be used to draw conclusions about an individual's health condition [10,11]. In fact, it was shown that the activity could also be used to distinguish different types of cancer [12].

Measuring and analyzing the dynamics of proteolytic processes often relies on array-based systems (see for example [13]), which have a high sensitivity. But this comes at the expense of high specificity to a single proteolytic process. In contrast to this we present a new method, that is able to reconstruct a proteolytic process and its kinetic parameters from mass spectrometry time series data. Mass spectrometry has become an essential tool in the field of proteomics [14] and can be used for the analysis of complex biochemical events, such as proteolytic processes (for a good overview see [15]).

The basic idea in these experiments is to incubate peptides (or proteins) with one ore many proteases and to generate mass spectra in every chosen time step that reflect snapshots of the proteolytic process. Figure 1 shows two snapshot spectra of such an incubation experiment after seven and 24 hours of incubation. One can clearly see how a large peptide of about 2680 Da (represented by the large peak to the right in the upper spectrum) is degraded into smaller fragments (represented by large peaks to the left in the lower spectrum). The fragments (represented by peaks in the lower spectrum) are generated by two different degradation reactions: exo- and endoproteolytic cleavage. During an exoproteolytic reaction a single amino acid is removed from one of the free termini of a molecule, while in an endoproteolytic

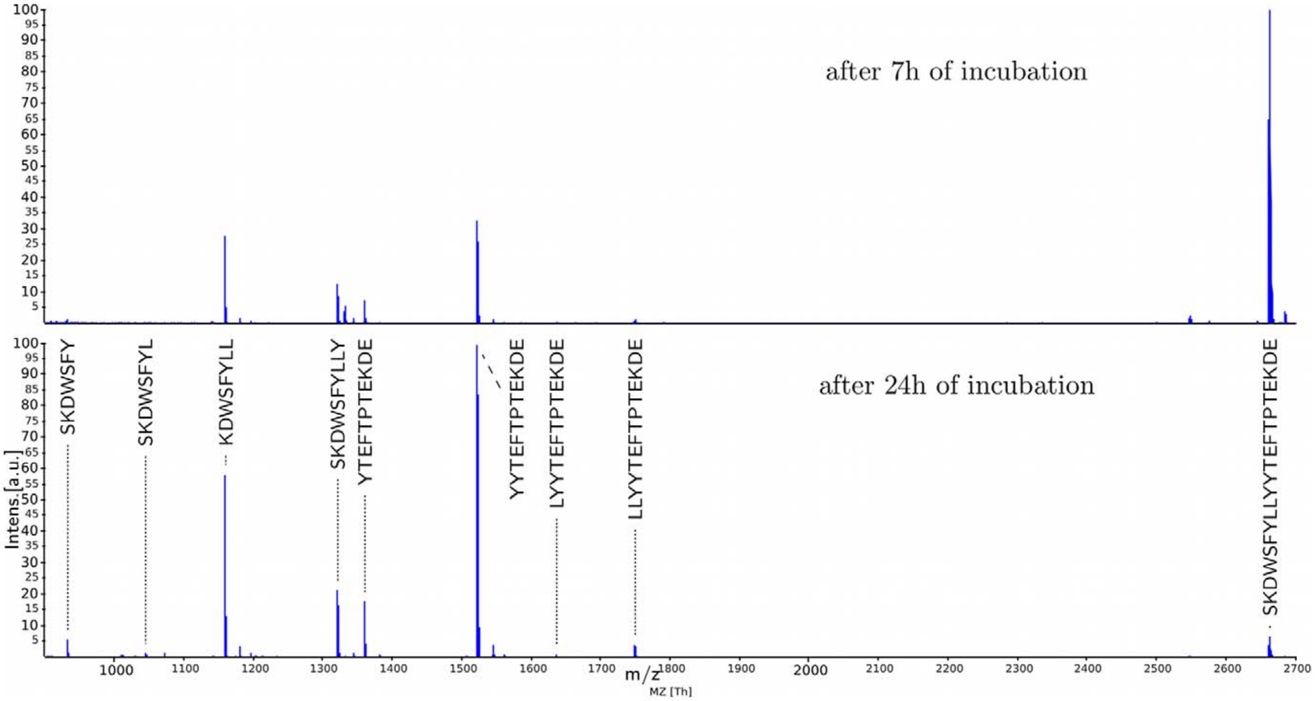


Figure 1. Degradation of a beta-2-microglobulin fragment observed via Mass Spectrometry. Mass spectra generated during the degradation of a beta-2-microglobulin fragment by a mixture of urine proteins after 7 (upper) and 24 (lower) hours of incubation. Intensity is given in percent of maximal peak intensity. In the lower spectrum all fragments were annotated that could be verified by MS/MS identifications. Details of data acquisition and sample preprocessing are given in the results section. All raw spectra of this time series are shown in the Supporting Information (Figure S2).
doi:10.1371/journal.pone.0040656.g001

reaction the targeted molecule is cleaved at a position between the N- and C-terminus.

Compared to similar state-of-the art approaches by Yi et al. [16] and Kluge et al. [17] our contribution to the field is the inclusion of endoproteolytic degradation while using differential equations to model the full dynamics of the underlying process. In contrast to that the work of Yi et al. was only focused on a single proteolytic process (the degradation of fibrinopeptide A) and the method presented by Kluge et al. only considered exoproteolytic reactions and used a statistical model to describe the dynamics, independent of the degraded molecules.

An example result of our method is shown in Figure 2. Here we show a *degradation graph*, a data structure we will introduce later in detail, illustrating how a small peptide is degraded during several steps into smaller fragments. The kinetic constants of the individual reactions are omitted for the sake of clarity. The workflow of our method - which will be described in more detail in the remaining part of this paper - is as follows:

- Perform incubation experiment and generate mass spectra at every chosen time point.
- Create an initial degradation graph from the time series.
- Optimize the degradation graph structure by removing unlikely reactions and peptides and estimate the kinetic parameters of the generated model.

The results of our method are the sequences of all intermediate peptides, the proteolytic reactions that connect those peptides, as well as the dynamics of all proteolytic events.

In the result section we intensively test our approach on multiple simulated data sets with varying conditions. It shows a good performance in recovering the original structure as well as the underlying reaction rates. We further prove the applicability of our method to a real data set using a time series of a peptide incubated with an unknown mixture of urine proteins.

Methods

Biochemical processes (such as proteolysis) can be described by ordinary differential equations (ODEs). This allows to simulate and analyze a process and thus to draw conclusions about its properties, such as steady-states or changes in concentration of its constituents over time. A simple example for such a system is Tyson's cell cycle model [18]. To visualize these ODE systems oftentimes graphs are used, where nodes are the reactants and edges between them are the reactions. Note that both representations (ODE and graph) are equivalent. For modeling and visualizing proteolytic processes Kluge et al. introduced the *cleavage graph* [17] which they used to model exoproteolytic cleavage reactions. In the following we will extend this concept to also include endoproteolytic reactions. We call the resulting data structure *degradation graph* since it can be used to model all degradation reactions of a proteolytic process and also allows a convenient and comprehensible visualization.

Degradation Graph

A proteolytic process where single or multiple peptides are generated by cutting peptides into smaller fragments can be modeled as a graph $G=(V,E)$.

The nodes V correspond to the degraded and generated peptides and the edges E to the proteolytic reactions. Since proteolysis is an irreversible reaction under physiological

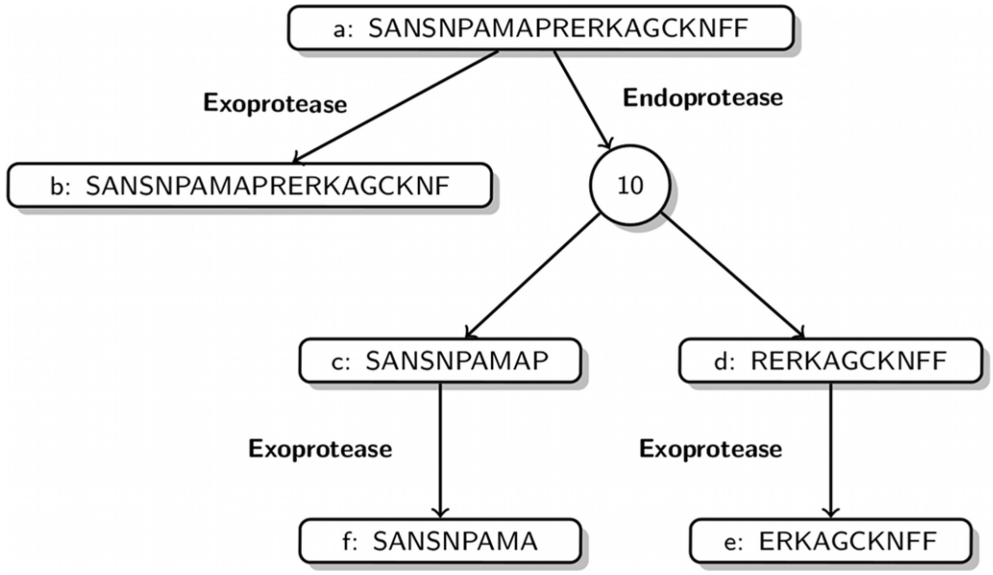


Figure 2. Complex proteolytic reaction visualized as graph. Example protease system acting on a single peptide (SANSNPAMAPRERKAGCKNFF) and the resulting degradation products. The shown reactions are all artificial.
doi:10.1371/journal.pone.0040656.g002

conditions the edges in the graph are directed from the degraded to the generated peptides.

As mentioned above, one can distinguish two types of proteolytic reactions, exoproteolytic reactions, where a single amino acid is removed from one of the free termini of the peptide, and endoproteolytic reactions, where the targeted peptide is cleaved at a position between the N- and C-terminus. For exoproteolytic reactions we connect two nodes with a directed edge from node *u* to *v* if we can obtain the amino acid sequence of *v* by subtracting a single amino acid from the beginning or the end of the amino acid sequence of *u*. For endoproteolytic reactions this is not that easy. Since we need to connect three nodes (the peptide that is targeted *u* and the two resulting fragments *v*,*w*) we need to break the idea of one reaction equals one edge in the graph. To ensure that we still associate the reaction with single edge, we introduce pseudo-nodes u_c, that represent the endoproteolytic process of cutting the peptide *u* at a specific position *c*. The pseudo-nodes can also be seen as representation of the endoprotease that cuts the peptide *u* at position *c*. We can now connect *u* to u_c and associate all reaction specific information (e.g., reaction rate) with this single edge. We further connect u_c to *v* and *w* with so called pseudo-edges.

Both reaction types are separately shown in Figure 3. An example with real peptide sequences and both reaction types is shown in Figure 2.

Constructing the Graph from Mass Spectrometry Data

In the previous section we defined the degradation graph and its relation to proteolytic processes. Now we present an approach to construct this graph based on series of N mass spectra collected at different time points $t_1 \ldots t_N$ and a seed sequence S which we will also call base peptide from here on. Based on this input we try to identify signals in the mass spectra, that represent fragments of S produced by a proteolytic process. The seed sequence needs to be provided as input. It can for instance be the sequence of a known peptide probe that was incubated with an unknown mixture of proteases or a sequence taken from MS/MS identifications.

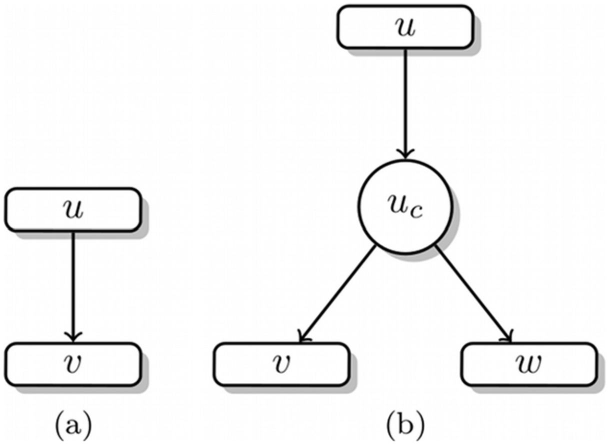


Figure 3. Representation of the basic degradation graph structures. (a) Exoprotease reaction, (b) Endoprotease reaction. See Figure 2 for an example containing both reaction types.
doi:10.1371/journal.pone.0040656.g003

We shortly introduce some notation that eases the understanding of the following explanations. Given a node *v* in the degradation graph, $s(v)$ denotes the amino acid sequence of the peptide associated with the node *v*. The length of the amino acid sequence is given by $|s(v)|$. $s(v)[a,b]$ with $1 \leq a \leq b \leq |s(v)|$ is the subsequence of the amino acid sequence from position *a* to position *b*. $m(v)$ denotes the mass of the peptide associated with the node *v*. If we could identify a signal that corresponds to the peptide associated with *v*, we will denote it's intensity with $I_{m(v)}(t_i)$. The association between mass and intensity takes into account, that mass spectrometers measure only mass to charge ratios and therefore cannot distinguish peptides with equal mass. Therefore different peptides with equal mass can be associated to the same intensity value, without counting the signal twice in the later analysis. The set of all peptide masses in the graph is denoted by

M. We further introduce a queue of nodes L, which is empty at the beginning of the construction.

The construction of the graph is divided into two parts, verification and extension, which are executed on each of the input spectra. Before we can execute these steps, we need to initialize the degradation graph. This is done by adding a node for the seed sequence to the degradation graph. Afterwards we start with the verification step for the first spectrum recorded at time point t_0, followed by the extension step. This is repeated for each of the input spectra. The pseudocode for both parts is shown in the Supporting Information (Figure S1).

Verification. The first step is the verification of the degradation graph on the new spectrum. We therefore check for each node in the degradation graph whether we can find a signal that corresponds to this node in the spectrum. In general, we will identify signals by peptide mass fingerprinting [19]. Our approach is described in the Supporting Information (Text S1). Existing MS/MS identifications [20] are solely used for validation, since relying only MS/MS identifications during the construction phase of the algorithm would introduce a bias towards the used acquisition strategy. Each node v that could be identified in the spectrum is added to L and annotated with the observed intensity $I_{m(v)}(t_i)$.

Extension. The extension step is performed on the current spectrum as long as L is not empty. In each cycle a node u is removed from L and the following procedure is executed.

Given the node u, we start by removing the N- and C-terminal amino acid separately from $s(u)$ to simulate exoproteolytic degradation and search for the corresponding signals. If we find a signal we add the corresponding node v to the graph, annotate it with the signal intensity $I_{m(v)}(t_i)$, set it's sequence $s(v)$ to either $s(u)[2,|s(u)|]$ or $s(u)[1,|s(u)|-1]$, and connect the nodes u and v by an edge pointing from u to v. The generated node v is appended to the list L.

Subsequently we simulate the endoproteolytic reactions by splitting the sequence $s(u)$ in two parts at each position c with $2<c<|s(u)|-1$. If we can identify both fragments of such a split in the mass spectrum, we add a pseudo-node u_c, annotated with the sequence $s(u)$ and the cutting position c to the graph and connect it to the degraded node u. We then add nodes v and w for each of the fragments to the graph, annotate it with the corresponding signal intensities $(I_{m(v)}(t_i), I_{m(w)}(t_i))$, the sequences $(s(u)[1,c]$ and $s(u)[c+1,|s(u)|])$, and connect it to the pseudo-node u_c. The generated nodes v,w are appended to the list L.

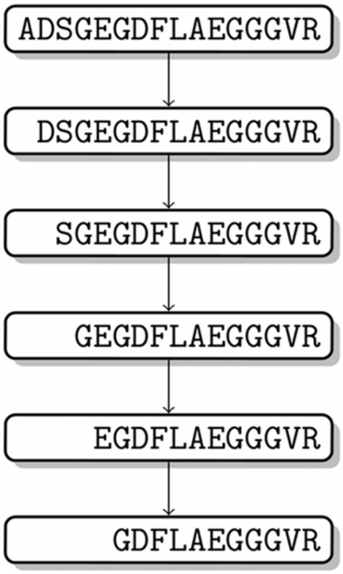


Figure 4. Degradation graph of the degradation of fibrinopeptide A (FPA) as reported in [16].
doi:10.1371/journal.pone.0040656.g004

Estimation of Kinetic Parameters

After we generated the model representing the proteolytic process, i.e., the degradation graph, the next task is to estimate the kinetic parameters of the underlying process. To achieve this we first generate a system of ordinary differential equations (ODE) based on a degradation graph as described in the following section. For this system we estimate the kinetic parameters based on the observed signal intensities.

Generating an ODE Model for the degradation graph. Following the ideas presented by Yi et al. [16] the mathematical model is derived by the law of mass action and each proteolytic reaction is modeled as a first-order reaction, i.e., the rate of the reaction depends on the concentration of only one reactant. In case of proteolytic reactions, this reactant is the protein or peptide that is degraded. We neglect side effects like saturation of the degradation products but incorporating these would be possible by an extension of the ODE system. We write the rate equations for an exoprotease reaction, where u is degraded to v as follows

$$\frac{dC_u(t)}{dt} = -k_{uv}C_u(t)$$

$$\frac{dC_v(t)}{dt} = k_{uv}C_u(t)$$

where $C_u(t)$ and $C_v(t)$ denote the concentration of peptide u and v at time t. k_{uv} is the kinetic rate constant for the reaction. Endoprotease reactions are represented in the same manner with the slight difference that we need to model both degraded products.

$$\frac{dC_u(t)}{dt} = -k_{uvw}C_u(t)$$

$$\frac{dC_v(t)}{dt} = k_{uvw}C_u(t)$$

$$\frac{dC_w(t)}{dt} = k_{uvw}C_u(t)$$

This transformation can be done for each reaction and each reactant in the degradation graph. As an example we transformed the degradation graph shown in Figure 2 into the following system of differential equations.

$$\frac{dC_a(t)}{dt} = -k_{ab}C_a(t) - k_{acd}C_a(t) \qquad \frac{dC_b(t)}{dt} = k_{ab}C_a(t)$$

$$\frac{dC_c(t)}{dt} = k_{acd}C_a(t) - k_{cf}C_c(t) \qquad \frac{dC_d(t)}{dt} = k_{acd}C_a(t) - k_{de}C_d(t)$$

$$\frac{dC_e(t)}{dt} = k_{de}C_d(t) \qquad \frac{dC_f(t)}{dt} = k_{cf}C_c(t)$$

Since the degradation process as well as the mass spectrometry measurements happen *ex-vivo*, the base peptide ($C_a(t)$ in the above example) has a fixed starting concentration and there will be no further production of the base peptide. In settings where this does not hold, one would need to explicitly model the generation of the base peptide into the equations (e.g., by a constant generation rate).

Transforming peptide concentrations to signal intensities. The presented ODE model is based on concentrations of peptides but with a mass spectrometer we can only observe intensities associated with a specific mass. The obvious question is what kind of relationship exists for a single peptide between its concentration and the intensity observed with the mass spectrometer. Moreover one cannot guarantee that two peptides with equal concentration will have the same intensity in the mass spectrometer.

Different studies [21,22] have shown that for a single peptide a linear relationship between intensity and concentration is a reasonable assumption. Based on this we introduced a linear transformation from the model concentrations to the predicted signal intensities.

$$\hat{I}_m(t) = f_i C_i(t),$$

where $\hat{I}_m(t)$ is the intensity associated with the mass m at time point t, m is the mass of the peptide i, f_i is a peptide specific factor, and $C_i(t)$ the concentration, computed by the model, for peptide i at time point t. Yi et al. [16] already used a similar transformation successfully in their study. This transformation implicitly solves also the second problem of comparability between two observed intensities. Since each observed intensity will be transformed individually into the common concentration domain, the resulting concentrations can be compared afterwards. This transformation can also be used to compensate for systematic effects that occur in each measurement, e.g., quantification errors or incomplete ionization.

Another problem is that it can happen that two or more different peptides have the same or a nearly identical mass. These isobaric peptides cannot be distinguished in a mass spectrum. We therefore transform them into a single intensity value. For every observed mass m, we compute a linear combination of all peptide concentrations, of peptides with a mass equal (or nearly equal) to m.

$$\hat{I}_m(t) = \sum_{i \in P(m)} f_i C_i(t),$$

where $P(m)$ is the set of all peptides i which have a mass of m.

Estimating reaction rates. To estimate kinetic parameters we first generated an ODE model based on a degradation graph as described above. We now need to find the optimal set of model parameters (k_i) as well as transformation parameters (f_i), so that the difference between the computed model intensities $\hat{I}$ and the observed intensities I is minimal. Following standard practice we use a weighted sum of least squares differences between observed and model intensities as an error measure.

$$\min \sum_{m \in M} \left(\sum_{i=0}^{N} \left(\frac{\left(\hat{I}_m(t_i) - I_m(t_i)\right)^2}{w(m,i)} \right) \right)$$

where M is the set of all observed masses, $I_m(t_i)$ is the intensity

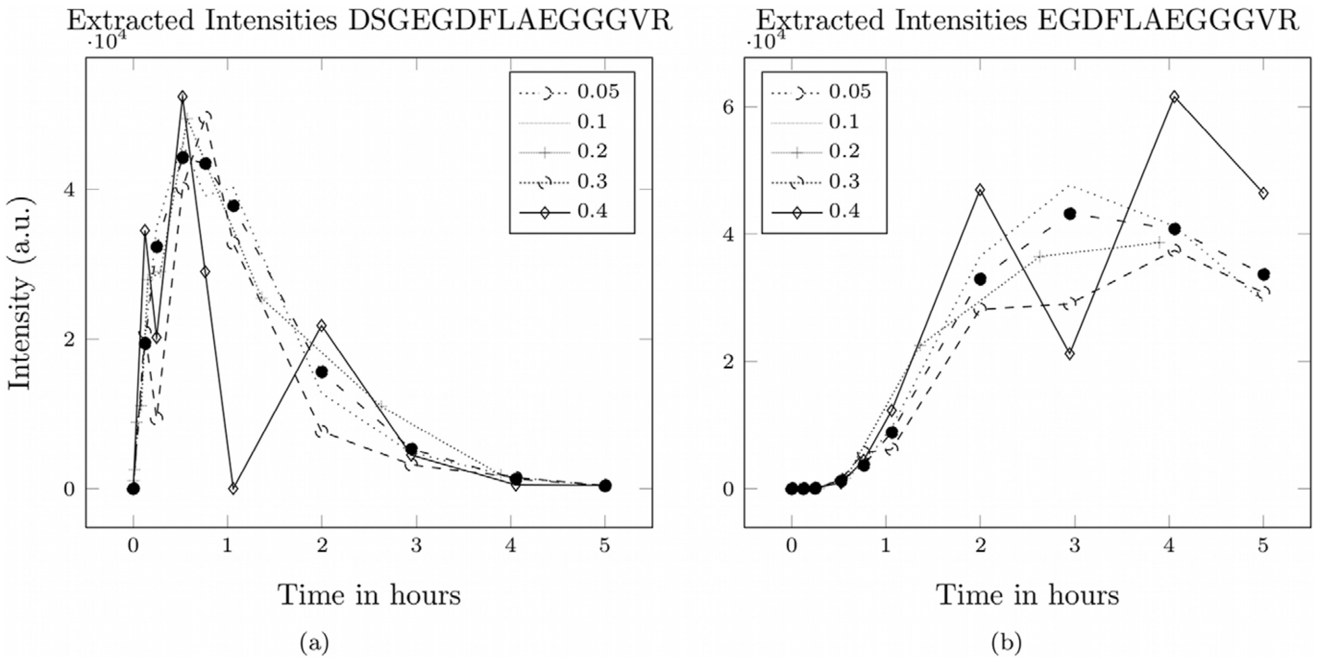


Figure 5. Effect of the different signal variability settings on the simulated signal intensities. Shown are the extracted signal intensities for two peptides (a) DSGEGDFLAEGGGVR (left) and (b) EGDFLAEGGGVR (right) of the fibrinopeptide A system shown in Figure 4 with increasing signal variability values.
doi:10.1371/journal.pone.0040656.g005

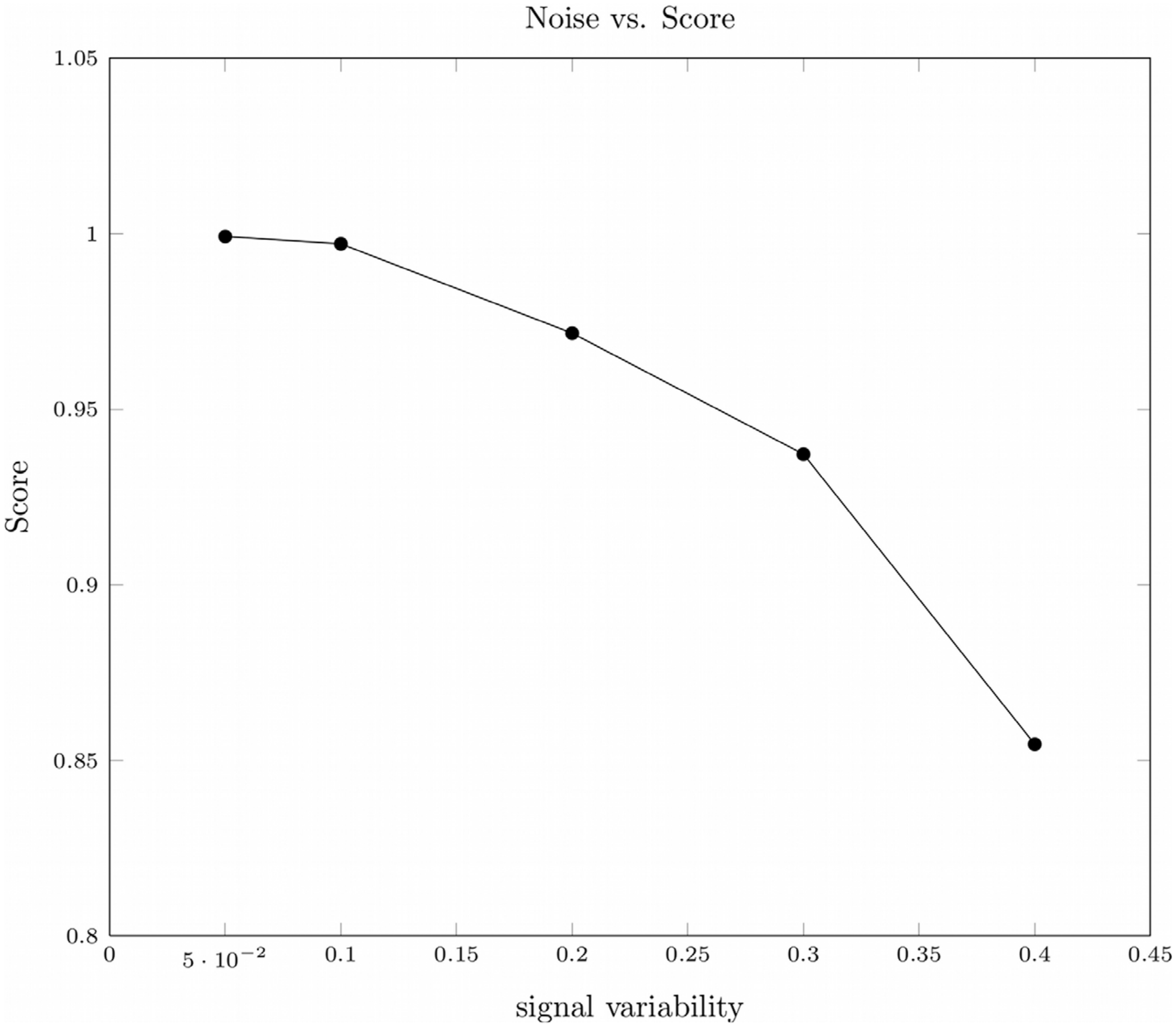


Figure 6. Effect of the signal variability on the score S. Effect of the variability of the signal with respect to the intensity on the score S computed by our method. Data was generated based on the fibrinopeptide A system shown in Figure 4.
doi:10.1371/journal.pone.0040656.g006

observed for mass m at time point t_i, $\hat{I}_m(t_i)$ is intensity predicted by ODE system for the mass m at time point t_i, and w is a weighting function. The weighting function can for instance be used to use relative instead of absolute deviations, i.e.,

$$w(m,i) = I_m(t_i).$$

This is used to reduce the effect of different intensities being on different orders of magnitude. This minimization problem can theoretically be solved by any available optimization technique. After testing different available techniques we decided to use POEM, a Matlab-based version of BioPARKIN [23,24], to estimate the model parameters as well as the transformation parameters. We further use POEM to estimate the initial concentration of the base peptide. POEM is based on damped Gauss-Newton techniques for solving the above optimization problem. Lack of robustness of damped Gauss-Newton techniques as observed often in model discrimination contexts, see [25], can be overcome by using dimension reduction in parameter space [26].

How to choose initial values. As the prior knowledge on the modeled system is very limited good initial values for the estimation of the model parameters are hard to find. We therefore chose the initial values based on the following scheme: For each node the edge (i.e., proteolytic reaction) is selected, which leads on the shortest path to the root node. For the corresponding reaction rate (k_i) we assign an initial value of 1.0. For all other incoming reactions the initial value is set to a value of 1×10^{-6}. All transformation parameters (f_i) are set to 1.0.

Evaluation and Optimization of the Degradation Graph Structure

The above presented approach to construct the degradation graph is greedy, i.e., it assumes that every signal in a spectrum that could match a subsequence of the base peptide is part of the

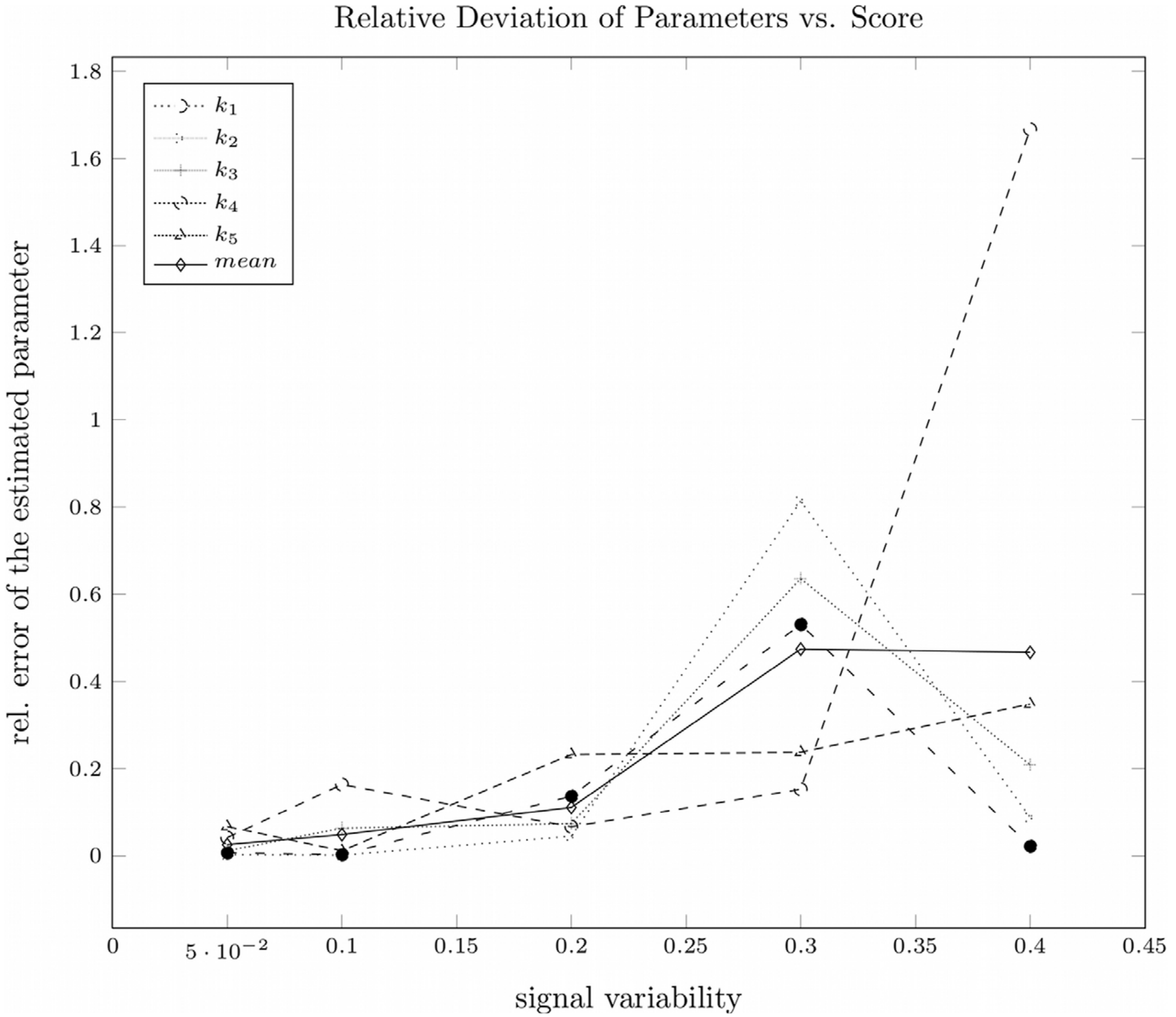


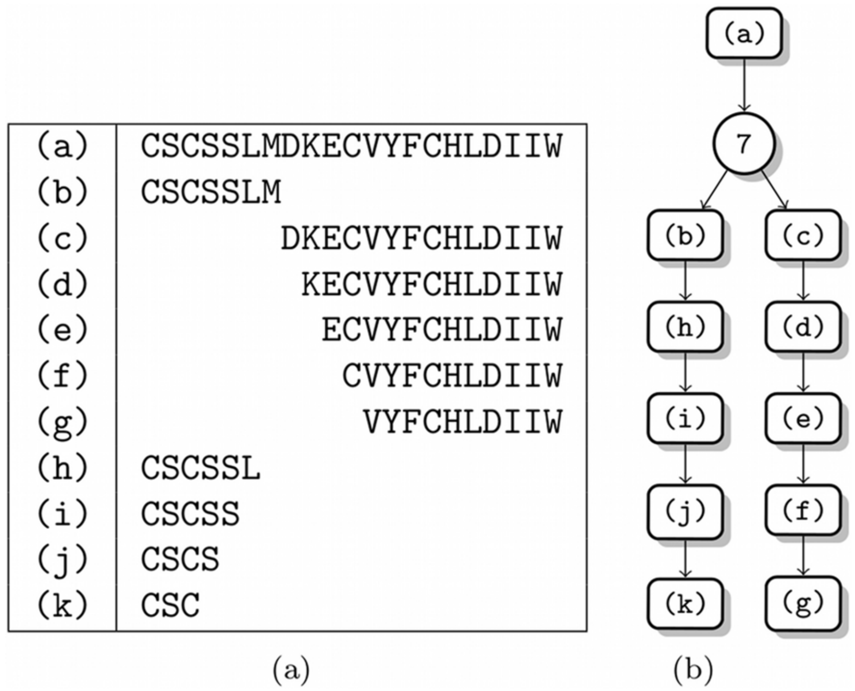
Figure 7. Effect of the signal variability on the the relative error of the estimated parameters. The quality is given in terms of the relative deviation of the estimated from the real parameter $\left(\frac{|p_{real} - p_{est}|}{p_{real}}\right)$. Data was generated based on the fibrinopeptide A system shown in Figure 4. The reaction parameters are numbered in the order of degradation (e.g., FPA → FPA-1 $= k_1$) shown in Figure 4.
doi:10.1371/journal.pone.0040656.g007

proteolytic process and that every possible reaction occurred. This assumption is not always true. The signals could also originate from peptides with equal or at least similar masses as we have already seen in the previous section. But these peptides do not necessarily take part in the proteolytic reactions, that we want to model. We will call such peptides *decoy* peptides. Alternatively we may have multiple reactions to explain the formation of a peptide where only one is true. Hence, the degradation graph may contain peptides or reactions that did not occur in the actual underlying proteolytic process. To account for this we present a method to rank different subgraphs of an initial degradation graph with respect to their ability to explain observed data. Followed by a heuristic approach to construct a series of smaller models from the initially generated degradation graph without the need to compute every possible subgraph.

Evaluating different models. To find the degradation graph that optimally explains the observed data it is necessary to rank the different graphs. Here we describe a scoring scheme that can be used to rank the generated models.

To ease the following explanations we will introduce some further notation. Given a degradation graph G, a subgraph G' is defined as $G' = (V', E')$, where $V' \subseteq V$ and $E' \subseteq E$. We also require that G' is connected, i.e., for all pairs of nodes $u,v \in V'$ exists a path of length n in E'^n that connects u and v. The subgraph G' also defines $M' \subseteq M$ as the subset of all masses m and their associated intensities that are explained by the subgraph $M' = \{m(v), v \in V'\}$.

The proposed score consists of two components. The first score component S_C is the average Pearson correlation of the intensities predicted by the model (with estimated reaction parameters) and

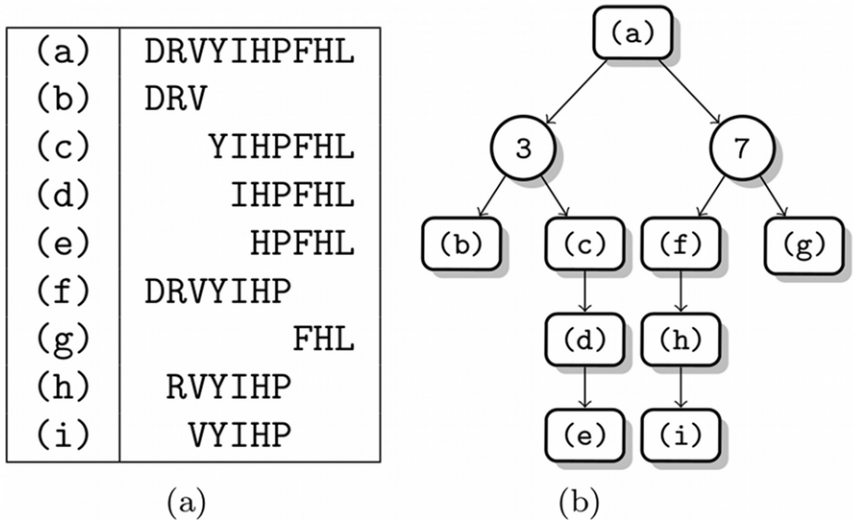


(a)	CSCSSLMDKECVYFCHLDIIW
(b)	CSCSSLM
(c)	DKECVYFCHLDIIW
(d)	KECVYFCHLDIIW
(e)	ECVYFCHLDIIW
(f)	CVYFCHLDIIW
(g)	VYFCHLDIIW
(h)	CSCSSL
(i)	CSCSS
(j)	CSCS
(k)	CSC

Figure 8. Endothelin-1 test system. Degradation of endothelin-1 by multiple artificial endo- and exoproteases. (a) The mapping of indices to sequences. (b) The degradation graph. package.
doi:10.1371/journal.pone.0040656.g008

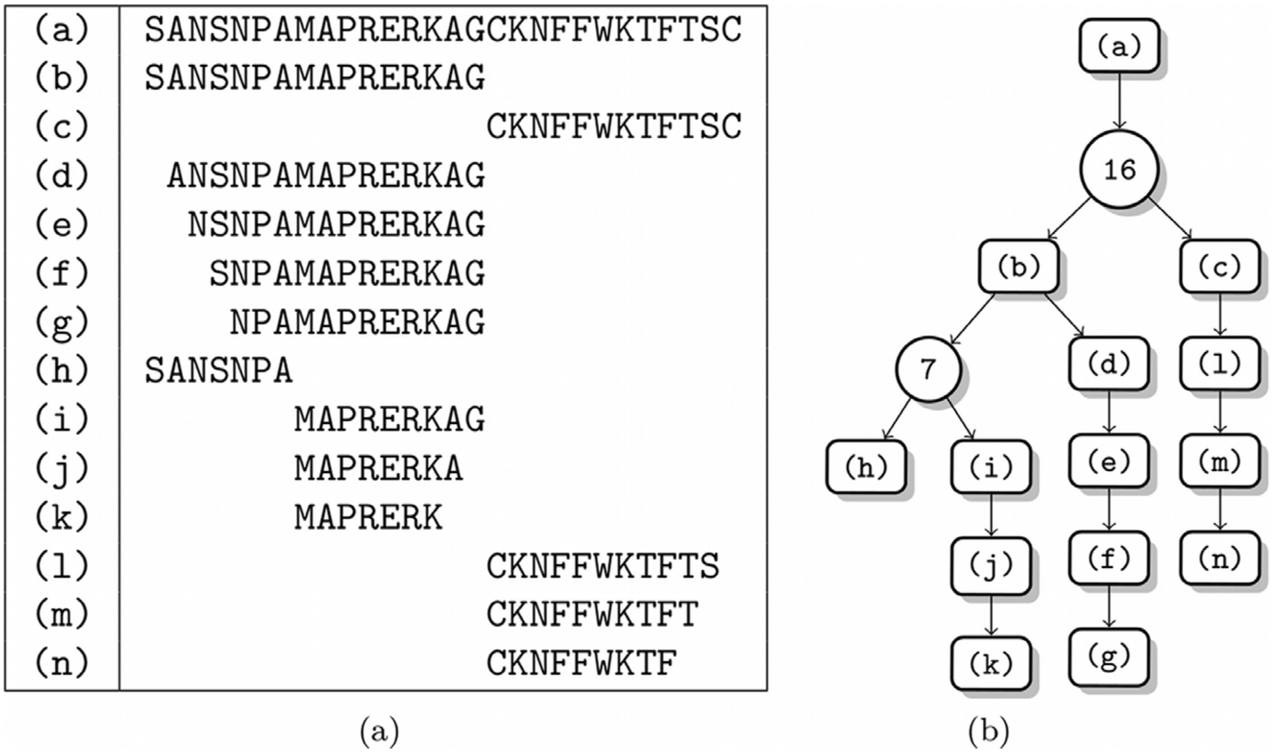
the actual observed data. This component should reflect the goodness of fit between the measured intensities and the computed model intensities. We compute for each explained mass $m \in M'$ the Pearson correlation r_m between the observed intensity values and the predicted values from the model.

$$r_m = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(\left(\hat{I}_m(t_i)-\bar{\hat{I}}\right)(I_m(t_i)-\bar{I})\right)}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{I}_m(t_i)-\bar{\hat{I}}\right)^2\frac{1}{N}\sum_{i=1}^{N}(I_m(t_i)-\bar{I})^2}}$$



(a)	DRVYIHPFHL
(b)	DRV
(c)	YIHPFHL
(d)	IHPFHL
(e)	HPFHL
(f)	DRVYIHP
(g)	FHL
(h)	RVYIHP
(i)	VYIHP

Figure 9. Angiotensin test system. Degradation of angiotensin by multiple artificial endo- and exoproteases. (a) The mapping of indices to sequences. (b) The degradation graph.
doi:10.1371/journal.pone.0040656.g009



(a)	SANSNPAMAPRERKAGCKNFFWKTFTSC
(b)	SANSNPAMAPRERKAG
(c)	CKNFFWKTFTSC
(d)	ANSNPAMAPRERKAG
(e)	NSNPAMAPRERKAG
(f)	SNPAMAPRERKAG
(g)	NPAMAPRERKAG
(h)	SANSNPA
(i)	MAPRERKAG
(j)	MAPRERKA
(k)	MAPRERK
(l)	CKNFFWKTFTS
(m)	CKNFFWKTFT
(n)	CKNFFWKTF

Figure 10. Somatostatin-28 test system. Degradation of somatostatin-28 by multiple artificial endo- and exoproteases. (a) The mapping of indices to sequences. (b) The degradation graph. package.
doi:10.1371/journal.pone.0040656.g010

.

$$\bar{\hat{I}} = \frac{1}{N}\sum_{i=1}^{N} \hat{I}_m(t_i), \bar{I} = \frac{1}{N}\sum_{i=1}^{N} I_m(t_i).$$

We then use the mean of all Pearson correlation values as measure for the goodness of fit.

$$S_C = \frac{1}{|M'|}\sum_{m \in M'} r_m.$$

The second component of the score S_V is the part of the standard deviation of the original degradation graph, that is conserved by the specific subgraph.

$$S_V = \frac{\sum_{m \in M'} s_m}{\sum_{m \in M'} s_m}$$

where s_m is the standard deviation of the signal corresponding to the mass m. S_V reflects the ability of the subgraph to explain the important parts of the originally collected signals.

To compute a single score S from these two components we build the weighted sum of both scores.

$$S = w_C S_C + w_V S_V$$

To determine good weights w_C and w_V we carried out several experiments on simulated data. A weight of $w_C = 0.9$ for the correlation score S_C and $w_V = 0.1$ for the variability score S_V showed the best separation of the correctly and wrongly identified models. For datasets with low quality (e.g., due to high amounts of noise or too few sampling points) weights of $w_C = 0.8$ and $w_V = 0.2$ have shown a good performance. For such datasets we expect a less reliable fit for the time series and therefore decreased the weighting factor for the quality of the fit.

Heuristic search for the optimal graph. Constructing all possible subgraphs, generate the associated ODE system, and estimating the corresponding reaction and transformation parameters is possible for small graphs. With increasing size in terms of number of nodes and reactions, estimating the reaction and transformation parameters for all subgraphs gets more computationally intensive. If we want to generate each possible combination of reactions we would get $2^{|E|}$ possible subgraphs. Even if we filter out some of the subgraphs (e.g., those who do not contain the root node or are not connected) we would still have to consider exponentially many subgraphs. For each of these subgraphs we would then need to derive the associated ODE system and estimate the reaction and transition parameters.

To speed up this procedure we present a heuristic approach. Preliminary tests have shown that the presented graph score improves, if the structure of the degradation graph gets closer to the original one. This can be explained based on the composition of the score. The first component reflects the goodness of fit between model and observed data. This should improve if we remove peptides and reactions, that do not belong to underlying process. The second component reflects the variability of the signals. If we remove only nodes that do not participate in the

Table 1. Parameter estimation error for the endothelin 1 system.

Parameter	p_{real}	p_{est}	$\lvert p_{real}-p_{est}\rvert$	$\frac{\lvert p_{real}-p_{est}\rvert}{p_{real}}$
k_{jk}	1.30	0.949	0.351	0.270
k_{ij}	1.90	2.496	0.596	0.314
k_{hi}	2.10	2.369	0.269	0.128
k_{bh}	1.05	0.955	0.095	0.091
k_{abc}	3.50	5.025	1.525	0.436
k_{fg}	2.30	1.351	0.949	0.414
k_{cd}	4.30	4.284	0.016	0.004
k_{ef}	0.30	0.380	0.080	0.265
k_{de}	2.10	2.015	0.085	0.040

Relative and absolute deviations of the estimated parameter values for the endothelin 1 system. The indices for the parameter names are taken from Figure 8. p_{real} denotes the parameter values used for the initial simulation and p_{est} the value estimated by the presented approach. The last two columns contain the absolute and the relative deviation of the estimated from the real parameter value.
doi:10.1371/journal.pone.0040656.t001

reaction, i.e., whose variability is low compared to the signals of peptides which are degraded and produced, this score component should still be near to the optimal value.

Based on the construction algorithm we know that the identified degradation graph is maximal in the sense that it contains all signals that were produced by the assumed process and possibly also parts that do not belong to the process. To find the optimal subgraph we start by removing all terminal reactions of the graph (i.e., reactions that produce at least one leaf) separately. For each of these subgraphs we estimate the kinetic parameters as described earlier. Subsequently we rate all subgraphs according to the criteria presented above. Then we take the best N models and again remove all leafs separately. We continue with this procedure as long as we can find at least one graph whose score is under the top N of all so far computed subgraphs and that was not trimmed in a previous iteration.

With this approach we can drastically reduce the amount of parameter optimizations that need to be carried out by still finding the originally embedded graph.

Preliminary tests have shown that setting N to either 2 or 3 is sufficient to effectively bound the number of unnecessary model evaluations while still identifying the original degradation graph.

Run-time Considerations

The above presented combination of degradation graph construction, parameter estimation and structure optimization requires a considerable amount of time, if the initial degradation graph is large. Therefore we now describe an approximation of the running time in the worst case. The run time of the initial degradation graph construction is determined by the number of verifications needed. Under the assumption that we would construct the complete degradation graph, i.e., all peptides are degraded in every possible way, one would create a degradation graph, which contains all possible substrings of the initial peptide sequence. Since we would need to verify each of this substrings once, the running time is in the worst case bounded by the maximal number of possible substrings of the initial peptide sequence. Given a seed sequence of length n we can construct at most $\frac{n(n-1)}{2}$ possible fragments, which could be checked in the spectrum. If we now analyze N spectra we will have at most $N\left(\frac{n(n-1)}{2}\right)$ verifications.

Table 2. Parameter estimation error for the angiotensin system.

Parameter	p_{real}	p_{est}	$\lvert p_{real}-p_{est}\rvert$	$\frac{\lvert p_{real}-p_{est}\rvert}{p_{real}}$
k_{fh}	0.50	0.498	0.002	0.004
k_{abc}	3.20	3.733	0.533	0.167
k_{afg}	1.80	2.226	0.426	0.236
k_{de}	1.05	1.111	0.061	0.058
k_{hi}	1.30	1.225	0.076	0.058
k_{cd}	1.50	1.320	0.180	0.120

Relative and absolute deviations of the estimated parameter values for the angiotensin system. The indices for the parameter names are taken from Figure 9. p_{real} denotes the parameter values used for the initial simulation and p_{est} the value estimated by the presented approach. The last two columns contain the absolute and the relative deviation of the estimated from the real parameter value.
doi:10.1371/journal.pone.0040656.t002

The complexity of the parameter estimation procedure can be approximated by $2N|E|^3$, where N is the number of time points, i.e., the number of evaluated mass spectra, and $|E|$ is the number of unknown parameters, i.e., the number of edges in the graph minus the number of edges connecting pseudo- and real nodes. Given this the time required for the parameter estimation will decrease with the subgraphs getting smaller. Under the assumption that even the proposed heuristic could require the computation of each subgraph, we would need to trigger $2^{|E|}$ optimizations in the worst case.

Results

To evaluate the ability of the presented approach we have to consider two parts: (1) reconstruction of the correct sequence of proteolytic events and (2) estimation of the reaction rates. The influence of different parameters like the complexity of the degradation graph or the variations of the signals in the mass spectrometer have to be assessed. This can only be done if enough data is available in terms of number of samples in varying quality. Both is not always given.

To overcome this problem we designed a series of simulated mass spectrometry data sets. The mass spectra were simulated using the software MSSimulator [27], a comprehensive simulator for mass spectrometry data. A detailed description of the software is given by Bielow et al. [27]. MSSimulator generates mass spectra based on a set of amino acid sequences and a configuration file, which contains all parameters necessary for the simulation, like ionization type or instrument resolution. In the following experiments all configuration parameters are hold fix, expect the signal variability. The signal variability is an intensity dependent deviation of the signal intensity of a single peptide signal, i.e., if we set an intensity noise value of 10% of the total signal intensity (area under the curve of the simulated peak) will vary with a standard deviation of 10% of the original signal intensity.

The time series for the simulated proteolytic process is generated based on the associated ODE system. The produced peptide concentrations are combined with the peptide sequences and are then put into MSSimulator.

All input and configuration files can be found in the Supporting Information (File S1, File S2). All generated mass spectra are post-

Table 3. Parameter estimation error for the somatostatin 28 system.

Parameter	p_{real}	p_{est}	$\lvert p_{real}-p_{est}\rvert$	$\frac{\lvert p_{real}-p_{est}\rvert}{p_{real}}$
k_{de}	0.70	0.719	0.019	0.027
k_{mn}	2.80	3.125	0.325	0.116
k_{lm}	1.20	1.145	0.055	0.046
k_{cl}	3.10	3.312	0.212	0.068
k_{jk}	2.40	1.998	0.402	0.167
k_{ij}	1.60	1.951	0.351	0.219
k_{ef}	1.24	2.032	0.792	0.639
k_{bd}	3.20	2.648	0.552	0.172
k_{bhi}	3.40	2.158	1.242	0.365
k_{abc}	4.3	3.760	0.540	0.126
k_{fg}	2.54	0.940	1.600	0.630

Relative and absolute deviations of the estimated parameter values for the somatostatin 28 system. The indices for the parameter names are taken from Figure 10. p_{real} denotes the parameter values used for the initial simulation and p_{est} the value estimated by the presented approach. The last two columns contain the absolute and the relative deviation of the estimated from the real parameter value.

doi:10.1371/journal.pone.0040656.t003

processed by the OpenMS PeakPicker [28] to transform the raw spectra into manageable pairs of mass-to-charge ratio and intensity.

We evaluate our approach on four different simulated models and one real data set. The first one is the degradation of fibrinopeptide A presented in [16], which is used to show the general performance under varying noise conditions. The later three are artificial systems constructed to show the applicability of the method to complex proteolytic processes. The real data set is a series of mass spectra collected during the incubation of a peptide probe with urine proteins.

Study 1 (Simulated Data): Validation Using the *ex vivo* Degradation of Fibrinopeptide A (FPA)

To demonstrate that our approach is able to recover the correct sequence of proteolytic events, i.e., the degradation graph and the corresponding kinetic parameters, we simulated a data set based on the fibrinopeptide A (FPA) (Swiss-Prot:P02671[20–35]) degradation, as described in [16]. It consists of a series of exoproteolytic cuts at the N-terminus of FPA. The corresponding degradation graph is shown in Figure 4. For the proteolytic reactions we used a slightly modified version of the kinetic parameters as published in [16]. The modified parameters can be found in the ODE formulation of the Supporting Information (File S1).

The proteolytic system was simulated over a time of 5 *h*. We generated 10 sampling points for the time series, five during the first hour of the incubation and the other five distributed equally over the remaining 4 hours. For these 10 time points we generated five sets of mass spectra with increasing signal variability of 5, 10, 20, 30, and 40% of the original signal intensity. The impact of the signal variability on the time course of the peptide intensities is shown in Figure 5. Thereafter we applied our new method to estimate the model structure as well as the kinetic parameters for each of the five time series.

Our method succeeded to reconstruct the original degradation graph as it is shown in Figure 4. The scores computed for the reconstructed systems show a clear dependency on the noise added during the mass spectra simulation (see Figure 6). The relative error for the individual parameters of the system in relation to noise on the simulated data is shown in Figure 7. These experiments show that even in the presence of extensive noise a valid reconstruction of the original process is possible. Also the estimated parameter values have an acceptable agreement with the original parameters. With a signal variability of 30% the quality of the estimated parameters starts to decrease drastically. This could possibly be mitigated by increasing the number of sampling points.

Study 2 (Simulated Data): Complex Degradation of Human Plasma Peptides

To test our method in a complex setting where also endoproteolytic reactions occur, we simulated the degradation of several human plasma peptides (and peptide fragments) by multiple artificial endo- and exoproteases. The targeted peptides were fragments of endothelin 1 (Swiss-Prot:P05305[53–73]), angiotensin (Swiss-Prot:P01019[34–43]), and somatostatin-28 (Swiss-Prot:P61278[89–116]). The full set of reactions and the corresponding peptide sequences are shown in Figures 8, 9, and 10.

All three systems were again simulated over a time of 5 *h*. We generated 15 sampling points from the time series. More sampling points were generated in the first hour of each time series, since during this time the systems change most. For all time points we generated mass spectra with a signal variability of 20%. During the mass spectrometry simulation of the systems we added decoy peptides that have masses similar to possible fragments of the base peptides. Therefore we also applied our method to iteratively optimize the structure of the degradation graph.

Our method generally succeeded to reconstruct the originally simulated degradation graphs. In case of the angiotensin system the peptide (e) was misinterpreted as IHPFH. Since both terminal amino acids of its predecessor (Leucin and Isoleucin) have equal mass they cannot be distinguished by the mass spectrometer hence both solutions are equally good.

For all three systems the estimated parameters in comparison to the original parameters are shown in Tables 1, 2, and 3. In general, the recovered parameters are quite well estimated. The average relative deviation between the estimated and the real parameters is between 10 and 20% for the different experiments. It can be observed that the largest errors occur towards the end of the degradation process (e.g., k_{ef} for the somatostatin 28 system). This can be due to the late formation of the later products and with this the lack of enough data points to effectively estimate the reaction parameters. An extension of the sampling range beyond 5 *h* or an increased sampling rate could possibly solve this issue.

Figure 11 shows the extracted intensities of two characteristic somatostatin-28 fragments compared with predicted model intensities. As one can see the predicted model intensities and the simulated intensities show a good agreement in their dynamic behavior.

Study 3 (Real Data): Validation on MALDI Time Series Data

To demonstrate the applicability of our method to experimental data, we analyzed a data set where a fragment of beta-2-microglobulin (Swiss-Prot:P61769[77–97]) was incubated with different urine proteins. Manual inspection of the mass spectra combined with the analysis of MS/MS spectra lead to the assumption, that four endoproteolytic reactions at positions 7–10

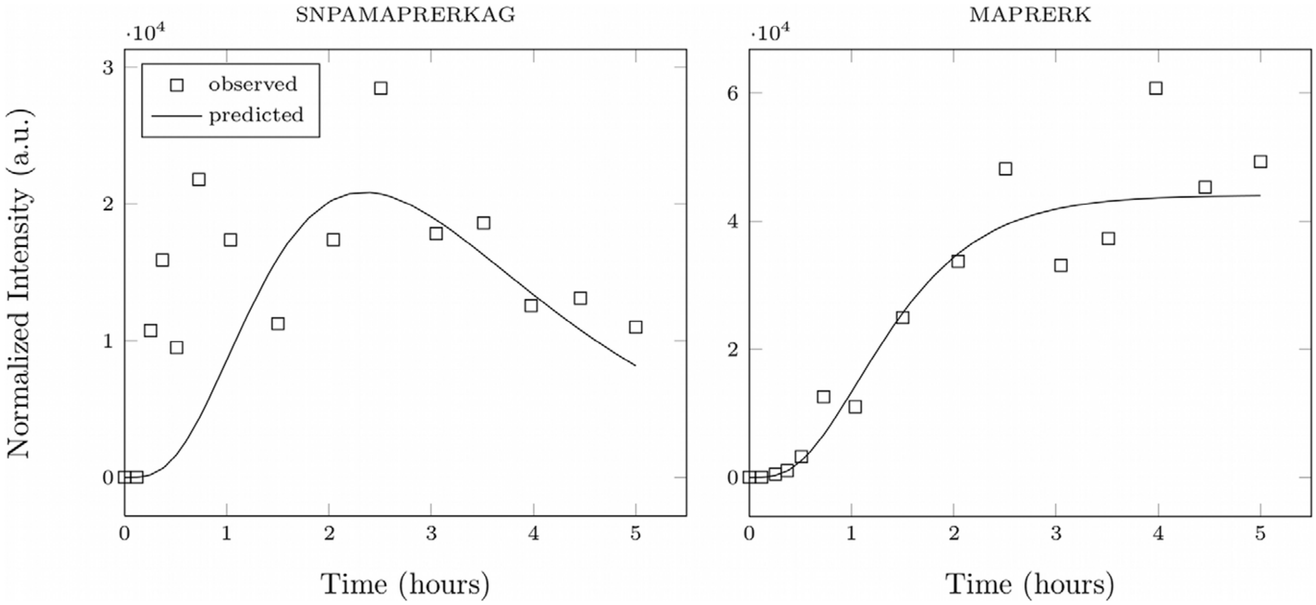
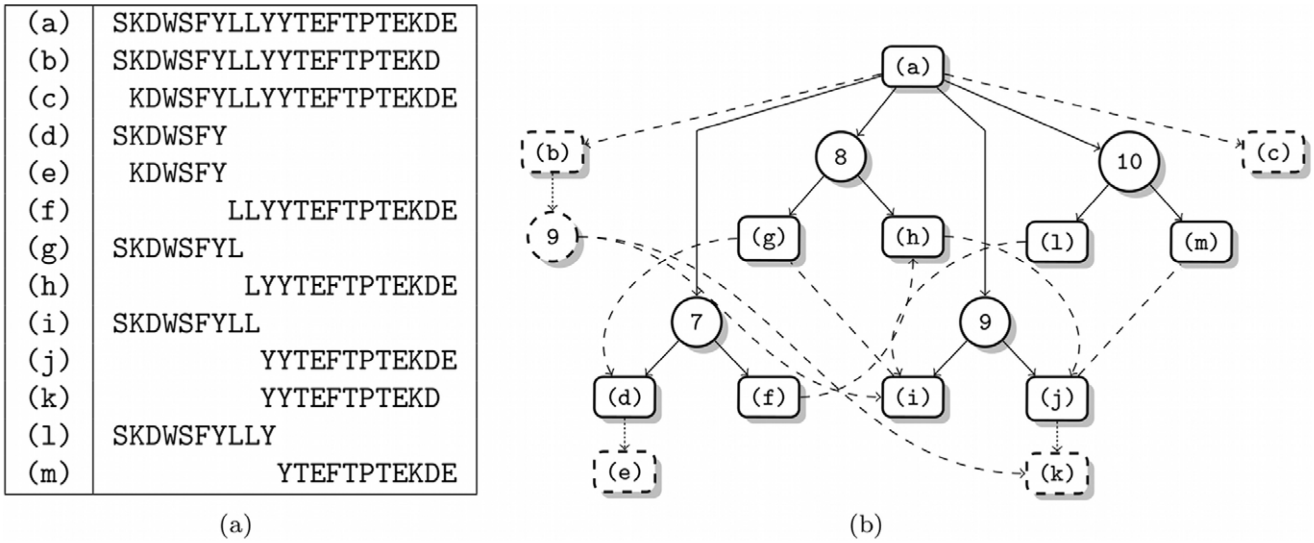


Figure 11. Intensity course for different fragments of the somatostatin-28 test system. Shown is the intensity course of two peptide fragments compared with the predicted model intensities for the best somatostatin-28 degradation graph.
doi:10.1371/journal.pone.0040656.g011

occurred. We applied the presented method to the dataset to validate this assumption.

The mass spectra for two time points ($t=7$ h and $t=24$ h) are shown in Figure 1. In the mass spectrum for $t=24$ h the peaks for the fragments generated by the four endoproteolytic reactions as well as the base peptide are annotated. A figure showing all mass spectra is included in the Supporting Information (Figure S2).

Data acquisition and preprocessing. For the immobilization of urine proteins from haemolytic urine of renal transplantation patients CNBr-activated Sepharosebeads® 6 MB were used. The Sepharosebeads® were incubated in 0.1 M hydrochloric acid (HCl) on a mixer (Horizontal Shaker, Rotator Drive STR4 Stuart Scientific, Redhill, England) for 30 min and washed with HPLC-grade water. The immobilization of urine proteins onto the Sepharosebeads® was done in coupling-buffer (100 mM $NaHCO_3$, 500 mM $NaCl$, pH 8.3) during an incubation period of 2 h on a mixer. Per preparation 50 μl urine and 30 μl Sepharosebeads® were used. After immobilization the Sepharosebeads® were washed with HPCL-grade water. Free binding capacities were saturated by over night incubation at 4°C in blocking-buffer (100 mM $NaHCO_3$, 500 mM $NaCl$, 0.2 M Glycin, pH 8.3). Afterwards the blocking-buffer was removed by washing with HPLC-grade water repeatedly.



(a)	SKDWSFYLLYYTEFTPTEKDE
(b)	SKDWSFYLLYYTEFTPTEKD
(c)	KDWSFYLLYYTEFTPTEKDE
(d)	SKDWSFY
(e)	KDWSFY
(f)	LLYYTEFTPTEKDE
(g)	SKDWSFYL
(h)	LYYTEFTPTEKDE
(i)	SKDWSFYLL
(j)	YYTEFTPTEKDE
(k)	YYTEFTPTEKD
(l)	SKDWSFYLLY
(m)	YTEFTPTEKDE

Figure 12. Initial degradation graph for the beta-2-microglobulin fragment estimated from real data. Shown is the degradation graph for the beta-2-microglobulin fragment which was initially estimated from a MALDI time series. (a) The mapping of indices to sequences. (b) The initial degradation graph. The dashed edges and nodes represent those reactions, that were not validated manually.
doi:10.1371/journal.pone.0040656.g012

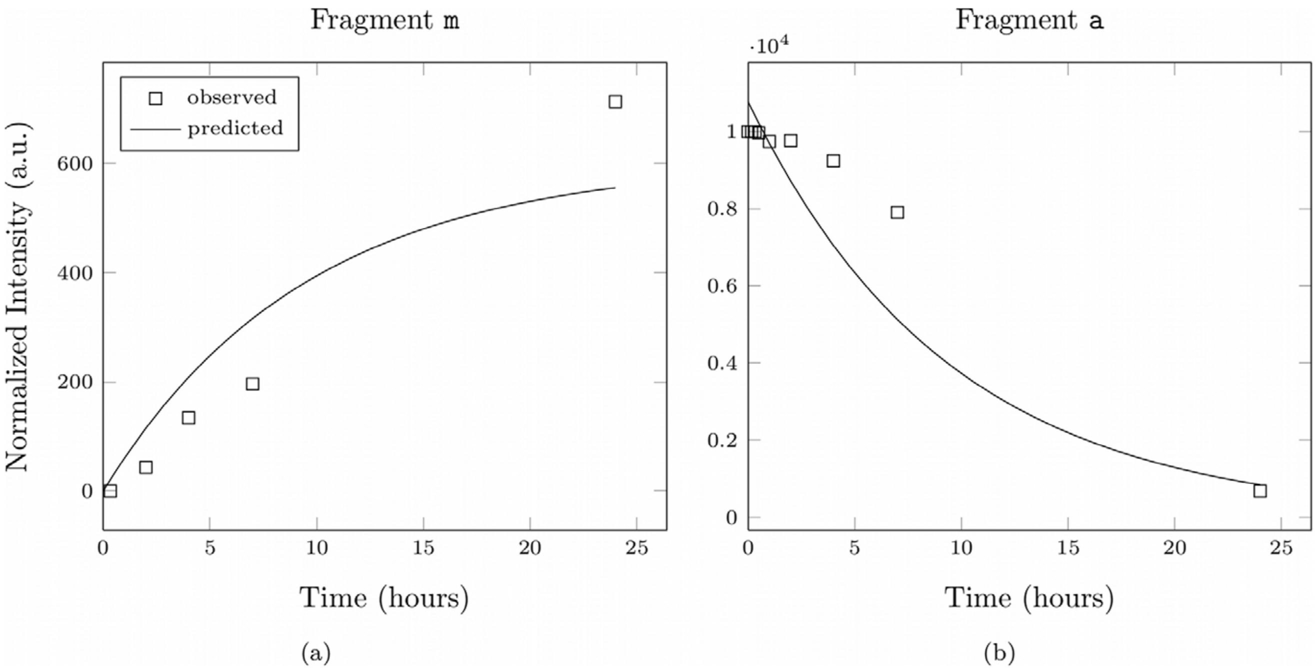


Figure 13. Intensity course for different fragments of the manually validated degradation graph. Intensity course for different fragments of the manually validated degradation graph. See text for more details.
doi:10.1371/journal.pone.0040656.g013

Incubation of immobilized urine proteins took place in sodium acetate buffer at pH 4.9 and was started by addition of the beta-2-microglobulin fragment to the immobilized proteins with a final concentration of 10^{-4}M in a reaction volume of 50 ml. At nine distinct time points aliquots were taken from the reaction mixture and diluted in a ratio of 1 : 10 in 0.2% (v/v) formic acid (Fluka/Sigma-Aldrich, Steinheim, Germany) for MALDI-TOF/TOF analysis on a 4700 Proteomics Analyzer (Applied Biosystems). The distinct time points were after 0, 10, 20, and 30 minutes, and 1, 2, 4, 7, and 24 hours.

All mass spectra were preprocessed as described in the simulation studies. To account for the variability of the overall intensity between different mass spectra we applied a customized normalization strategy to the intensities of the collected signals, which is described in detail in the Supporting Information (Text S2).

Results

The preprocessed spectra were analyzed by our method to identify the optimal degradation graph for the given mass spectra. The initially constructed degradation graph contains all four manually confirmed endoproteolytic cuts as well as four additionally not manually annotated exoproteolytic and one additional endoproteolytic cut. The complete degradation graph is shown in Figure 12. The unvalidated proteolytic reactions and fragments are represented as dashed nodes and lines.

It can further be seen from Figure 12 that the fragments generated by the validated endoproteolytic cuts are interconnected by exoproteolytic reactions. Although these reactions are possible, they are very unlikely and hence should be removed during the optimization. To reflect this the previously described selection of initial values was applied. Due to the lack of sampling points for the actual reactions, which took place between the last two sampling points, we have chosen the low quality weighting factors for this analysis ($w_C = 0.8$ and $w_V = 0.2$).

Optimizing the degradation graph structure results in a list of subgraphs ranked by their scores. The scores varied widely with the different generated structures. A figure showing the development of the score is included in the Supporting Information (Figure S3). Since the correct solution is unknown, we need to inspect the list and the different proposed solutions. As expected, based on the manual validation (see above), the degradation graph with the highest score contains the four endoproteolytic cuts at positions $7-10$. The unvalidated side reactions (see the dashed nodes and edges in Figure 12) were mostly removed, just two exoproteolytic reactions (fragment i to g and g to d) are still included, but have an estimated reaction rate of 1×10^{-6}. Although these reactions are still included in the degradation graph they have effectively no influence on the system and thereby can be neglected.

Figure 13 shows the observed and the predicted intensities for a subset of the peptide fragments of the degradation graph with the best score. The time courses of all peptides are shown in the Supporting Information (Figure S4). It can be seen that the measured intensities not always agree with the predicted intensity course, but they seem to show a comparable behavior. More time points especially in the time from 7 to 24 hours and an improved quantification (e.g., via a spiked in control sample) could further improve the results.

Conclusion

In this paper we presented a new method to model any proteolytic process as a degradation graph including an algorithm to construct the degradation graph based on mass spectrometry time series data. The degradation graph can easily be translated into a system of ordinary differential equations, which can be used to estimate the kinetic parameters of the proteolytic process. We further proposed an approach to optimize the initially constructed graph in the presence of decoy and overlapping signals. It is based on a score, that is used to rank the optimized and the original degradation graphs in their ability to explain the actually observed data. Using simulated data we have shown that our approach is able to compute good estimates for the kinetic parameters of the ODE systems even in the presence of noise and decoy signals. With a careful preparation of the samples using accepted standard operating procedures [29] the variability of the mass spectrometry data is below the observed boundary of the presented method. Applied to real data our approach reconstructed manually validated endoproteolytic reactions and removed unvalidated reactions and peptides from the graph.

We are aware of other biochemical approaches that give a much more robust and exact estimate for the reaction rates, but most of these methods rely on a much more time consuming measurement of the reactants and their concentration and, more importantly, often require prior knowledge of all reactants, which is not necessary for our method.

Applications for this method can be to identify and characterize unknown proteases and the estimated reaction kinetics can possibly be used to classify between different sample categories as it was done in [17]. With the ability to handle also false identifications the method can even be used in complex samples.

Future directions are an extensive validation of the proposed approach on real data. Another by now unsolved issue is the handling of unobserved peptides, i.e., peptides that participate in the reactions, but are not observable in the mass spectra. This can be due to different reasons e.g., the peptide cannot be ionized by the mass spectrometer or the degradation process is so fast, that the generated peptide is degraded before it can be measured. This problem can be handled by a modification of the construction algorithm for the degradation graph, as long as a downstream peptide is again observable. Also the handling of more than one seed sequence would be favorable. Finally a robust integration of MS/MS identifications into the method could further improve its performance. This could be done in two ways: One could use MS/MS identifications during the initial construction of the degradation graph in combination with the already used PMF approach, as an additional and more reliable way to identify possible fragments of the peptide probe. Furthermore one could integrate the MS/MS identification and its score into the scoring function by penalizing the removal of highly scored identifications.

The proposed score could also be improved in future development. The approach would benefit from a score that does not require specific scaling parameters for the different components. It would remove the step of optimizing the scaling parameters. First experiments using a scaled least-squares residual were carried out on simulated data sets. Those have shown promising but not yet comparable results.

The presented method is available on request. The whole approach is integrated into the proteomics.net platform [30]. The estimation procedure requires POEM which is available, for academic use, on request from the Computational Systems Biology Group, Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB) (http://www.zib.de/en/numerik/csb.html).

Supporting Information

Figure S1 Pseudocode for the degradation graph construction.
(PDF)

Figure S2 All mass spectra of the beta-2-microglobulin time series.
(PDF)

Figure S3 Ranked subgraphs of the beta-2-microglobulin analysis.
(PDF)

Figure S4 Time Courses of the confirmed beta-2-microglobulin fragments and the predicted dynamics.
(PDF)

Text S1 A simple peptide-mass-fingerprinting (PMF) strategy to extract intensity values from mass spectra.
(PDF)

Text S2 Description of the normalization method used for the beta-2-microglobulin data set.
(PDF)

File S1 Matlab code to generate the input data for MSSimulator to simulate the artificial models.
(ZIP)

File S2 Configuration files for the post-processing of the simulated and real data sets.
(ZIP)

Acknowledgments

The authors would like to thank Susanna Röblitz from the Computational Systems Biology Group, Konrad-Zuse-Zentrum für Informationstechnik Berlin for her help with POEM and fruitful discussions about the optimal formulation of the estimation problem and Xiaoxia Zhao for her help in measuring some of the test data sets.

Author Contributions

Conceived and designed the experiments: TC KR HS SA. Performed the experiments: SA DH. Analyzed the data: SA. Contributed reagents/materials/analysis tools: DH HS CS. Wrote the paper: SA. Provided background for the parameter estimation problem and the associated model discrimination problem: CS. Advised the mass spectrometry analysis: TC KR HS.

References

1. Thornberry NA, Lazebnik Y (1998) Caspases: Enemies Within. Science 281: 1312–1316.
2. Walsh PN, Ahmad SS (2002) Proteases in blood clotting. Essays Biochem 38: 95–111.
3. Alao JP (2007) The regulation of cyclin D1 degradation: roles in cancer development and the potential for therapeutic invention. Mol Cancer 6: 24.
4. Masaki T, Matsuoka H, Sugiyama M, Abe N, Goto A, et al. (2001) Matrilysin (MMP-7) as a significant determinant of malignant potential of early invasive colorectal carcinomas. Br J Cancer 84: 1317–21.
5. Leeman MF, Curran S, Murray GI (2003) New insights into the roles of matrix metalloproteinases in colorectal cancer development and progression. J Pathol 201: 528–34.
6. Ludwig T (2005) Local proteolytic activity in tumor cell invasion and metastasis. Bioessays 27: 1181–91.
7. De Clercq E (2007) The design of drugs for HIV and HCV. Nat Rev Drug Discov 6: 1001–18.
8. Cleynen I, Jüni P, Bekkering GE, Nüesch E, Mendes CT, et al. (2011) Genetic evidence supporting the association of protease and protease inhibitor genes with inammatory bowel disease: a systematic review. PLoS ONE 6: e24106.
9. Seife C (1997) Blunting Nature's Swiss Army Knife. Science 277: 1602–1603.
10. Liotta LA, Petricoin EF (2006) Serum peptidome for cancer detection : spinning biologic trash into diagnostic gold. J Clin Invest 116: 26–30.
11. Peccerella T, Lukan N, Hofheinz R, Schadendorf D, Kostrezewa M, et al. (2010) Endoprotease profiling with double-tagged peptide substrates: a new diagnostic approach in oncology. Clin Chem 56: 272–80.
12. Villanueva J, Shaffer DR, Philip J, Chaparro CA, Erdjument-Bromage H, et al. (2006) Differential exoprotease activities confer tumor-specific serum peptidome patterns. J Clin Invest 116: 271–284.
13. Steinrücke P, Aldinger U, Hill O, Hillisch A, Basch R, et al. (2000) Design of helical proteins for real-time endoprotease assays. Anal Biochem 286: 26–34.
14. Aebersold R, Mann M (2003) Mass spectrometry-based proteomics. Nature 422: 198–207.

15. Schlüter H, Hildebrand D, Gallin C, Schulz A, Thiemann J, et al. (2008) Mass spectrometry for monitoring protease reactions. Anal Bioanal Chem 392: 783–792.
16. Yi J, Liu Z, Craft D, O'Mullan P, Ju G, et al. (2008) Intrinsic peptidase activity causes a sequential multi-step reaction (SMSR) in digestion of human plasma peptides. J Proteome Res 7: 5112–5118.
17. Kluge B, Gambin A, Niemiro W (2009) Modeling exopeptidase activity from LC-MS data. J Comput Biol 16: 395–406.
18. Tyson JJ (1991) Modeling the cell division cycle: cdc2 and cyclin interactions. Proc Natl Acad Sci U S A 88: 7328–7332.
19. Pappin DJ, Hojrup P, Bleasby AJ (1993) Rapid identification of proteins by peptide-mass fingerprinting. Curr Biol 3: 327–332.
20. Nesvizhskii AI (2007) Protein identification by tandem mass spectrometry and sequence database searching. Methods Mol Biol 367: 87–119.
21. Chelius D, Bondarenko PV (2002) Quantitative Profiling of Proteins in Complex Mixtures Using Liquid Chromatography and Mass Spectrometry. J Proteome Res 1: 317–323.
22. Wang W, Zhou H, Lin H, Roy S, Shaler TA, et al. (2003) Quantification of Proteins and Metabolites by Mass Spectrometry without Isotopic Labeling or Spiked Standards. Anal Chem 75: 4818–4826.
23. Dierkes T, Wade M, Nowak U, Röblitz S (2011) BioPARKIN - biology-related parameter identifi- cation in large kinetic networks. Technical Report 11–15, ZIB, Takustr.7, 14195 Berlin.
24. Deuhard P (2004) Newton Methods for Nonlinear Problems. Affine Invariance and Adaptive Algorithms, volume 35 of *Springer Series in Computational Mathematics*. Springer, 1st edition, 424 pp.
25. Bock H, Körkel S, Kostina E, Schlöder J (2007) Robustness aspects in parameter estimation, optimal design of experiments and optimal control. In: Jger W, Rannacher R, Warnatz J, editors, Reactive Flows, Diffusion and Transport, Springer Berlin Heidelberg. 117–146.
26. Lorenz S, Diederichs E, Telgmann R, Schütte C (2007) Discrimination of dynamical system models for biological and chemical processes. J Comput Chem 28: 1384–1399.
27. Bielow C, Aiche S, Andreotti S, Reinert K (2011) MSSimulator: Simulation of mass spectrometry data. J Proteome Res 10: 2922–2929.
28. Lange E, Gröpl C, Reinert K, Kohlbacher O, Hildebrandt A (2006) High-accuracy peak picking of proteomics data using wavelet techniques. Pac Symp Biocomput: 243–254.
29. Baumann S, Ceglarek U, Fiedler GM, Lembcke J, Leichtle A, et al. (2005) Standardized approach to proteome profiling of human serum based on magnetic bead separation and matrix-assisted laser desorption/ionization time-of-ight mass spectrometry. Clin Chem 51: 973–80.
30. Conrad T, Leichtle A, Hagehülsmann A, Diederichs E, Baumann S, et al. (2006) Beating the Noise: New Statistical Methods for Detecting Signals in MALDI-TOF Spectra Below Noise Level. In: R Berthold M, Glen R, Fischer I, editors, Computational Life Sciences II, Springer Berlin/Heidelberg, volume 4216 of *Lecture Notes in Computer Science*. 119–128.