JOURNAL OF MASS SPECTROMETRY
*J. Mass Spectrom.* 2004; **39**: 188–192
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/jms.574

JMS

# Sequential exoproteolysis as a structural probe: a cautionary note

**Robert J. Beynon***

Protein Function Group, University of Liverpool, Crown Street, Liverpool L69 7ZJ

Received 21 July 2003; Accepted 2 October 2003

**In a recent paper, Villaneuva *et al.* (*J. Mass Spectrom.* 2002; 37: 974) described the use of exoproteases as probes of higher order structure in proteins. Their model assumes that the proteins are attacked sequentially from either the N-terminus or the C-terminus, depending on the type of exoprotease (aminopeptidase or carboxypeptidase) used. The products of this presumed exoproteolysis were then analysed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. The pattern of fragments obtained was mapped on to the primary sequence of the protein, and the exoproteolysis was interpreted as comprising a series of fast and slow phases, the rates of the different phases being directly related to the higher order structure of particular segments of the protein. Here, it is shown that this explanation is unlikely, that both kinetic and practical considerations suggest that alternative explanations for the data should be sought, and that exoproteolysis will perhaps not be as valuable as a conformational probe as the authors suggest. Copyright © 2004 John Wiley & Sons, Ltd.**

**KEYWORDS:** proteolysis; exoproteases; protein folding; mass spectrometry; limited proteolysis

## INTRODUCTION

The rapid emergence of a core technology in proteomics—that of peptide mass fingerprinting—means that we are becoming increasingly used to the analysis of limit peptides, i.e. the set of peptides that ensues after cleavage of every peptide bond in a protein compatible with the primary specificity of the proteinase. However, in most instances, to obtain limit peptides, it is desirable to eliminate conformational constraints on the rate of proteolysis by denaturation of the protein.

When proteins are fully denatured, all susceptible bonds have approximately similar rates of hydrolysis, and the routes of proteolysis are manifold.[1] In the progression of proteolysis to limit peptides, for any single substrate molecule, proteolysis is the sum of a sequence of reactions, and intermediates must be generated. However, the similarities of the individual rates of hydrolysis means that the rate of removal of partially digested intermediates is as high as their rate of formation, and large quantities of transient intermediates do not therefore accumulate. Further, each different pathway of proteolysis generates distinct intermediates, and the total mass of intermediate products is distributed over this peptide space such that discrete intermediates are rarely observed.[1] Hence, although individual protein molecules are digested at similar rates, they will generate different families of intermediates, depending on the sequence of cleavages.

*Correspondence to: Robert J. Beynon, Protein Function Group, University of Liverpool, Crown Street, Liverpool L69 7ZJ.
E-mail: r.beynon@liv.ac.uk
Contract/grant sponsor: Biotechnology and Biological Sciences Research Council.

By contrast, native protein structures can prove remarkably resistant to proteolysis.[2–5] The higher order structure of proteins means that many peptide bonds are inaccessible to the action of endoproteases or exoproteases, either because of steric inaccessibility or because of constraints on the segmental mobility of that region of the polypeptide chain.[6] In either circumstance, the outcome is that few, or sometimes none, of the peptide bonds can be cleaved by the proteinase. Proteolysis can also follow a single route of digestion (controlled by successive rates of hydrolysis) and intermediates are commonly generated.

Limited proteolysis has long been used a probe of protein structure[1–3] and there are many precedents for the isolation of proteolytically resistant fragment that are amenable to further analysis. By contrast, exopeptidases have rarely been used in this way. The restriction of the enzyme to sequential attack at the N-terminus (aminopeptidases) or C-terminus (carboxypeptidases) of the protein, coupled with the additional restrictions introduced by the differences in $k_{cat}/K_M$ for individual terminal amino acids means that the process can be inefficient and stall after rather limited truncation.

A recent paper by Villaneuva *et al.*[7] proposed the use of exoproteases as a probe of higher order structure and overall protein stability. Briefly, they exposed several native proteins to the exoproteases carboxypeptidase Y and leucine aminopeptidase, and monitored the appearance of digestion products by linear mode matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry. As the proteolysis progressed, the appearance of discrete fragments was interpreted as being attributed to the accumulation of protein that was relatively resistant

to proteolysis, generating 'stop sites' where exoproteolysis could not progress. However, this interpretation is not supported by consideration of the kinetics of sequential proteolysis. It is shown here that the behaviour they observed cannot be attributed to progressive exoproteolysis, and that alternative explanations are likely. Moreover, the fundamental nature of exoproteolytic attack on proteins means that exoproteolysis is of rather limited value for such studies.

## SEQUENTIAL PROTEOLYTIC REACTIONS

Most proteolytic reactions *in vitro* are conducted at very low concentrations of substrate and enzyme and show clear pseudo-first-order behaviour. Thus, a sequence of proteolytic reactions can be analysed as a series of first-order reactions. The simplest case comprises two reactions, described most simply as A → B → C, where $k_1$ is the first-order rate constant for the conversion A → B and $k_2$ is the first-order rate constant for the conversion B → C. The equation describing the appearance of the intermediate B is

$$[B] = k_1/(k_2 - k_1)[\exp(-k_1 t) - \exp(-k_2 t)]$$

It is apparent from this equation that both rate constants influence the outcome of the reaction, and that the persistence and amount of the proteolytic intermediate B depend on the relative values of the two rate constants. This is illustrated most clearly in the simulation in Fig. 1. If $k_2$ is zero, then B accumulates until all A has been converted to B. However, as $k_2$ becomes larger, the intermediate B shows an increasingly transient existence, attaining lower and lower values.

This simple scheme would apply equally well to an exoproteolytic attack in which there is one susceptible region and one resistant region. The complication introduced by multiple exoproteolytic clips is of little significance, since it has been assumed that unstructured regions of a protein undergo rapid attack by such enzymes, until a 'stop site' is reached.[7] Thus, endoproteolytic or exoproteolytic reactions can be used to generate transient or stable intermediates.

However, more complex schemes create problems caused by the alternation of slow and fast proteolytic steps such as those proposed by the authors. In particular, there are difficulties in constructing systems where there are more than one proteolytically resistant regions, interspersed with susceptible regions (Fig. 2). In qualitative terms, the first slow proteolytic step controls the formation of the product that in turn is susceptible to proteolysis. Although the rate constant for this fast step is high, the flux through the step can only be low, because the rate of appearance of substrate is correspondingly low, and as soon as substrates are generated, they are converted to product. The first slow step effectively becomes the rate-limiting step for all subsequent products, irrespective of the relative rates of formation. An effective analogy is provided by a series of pails placed one above the other. The first pail (the pool of starting material) is

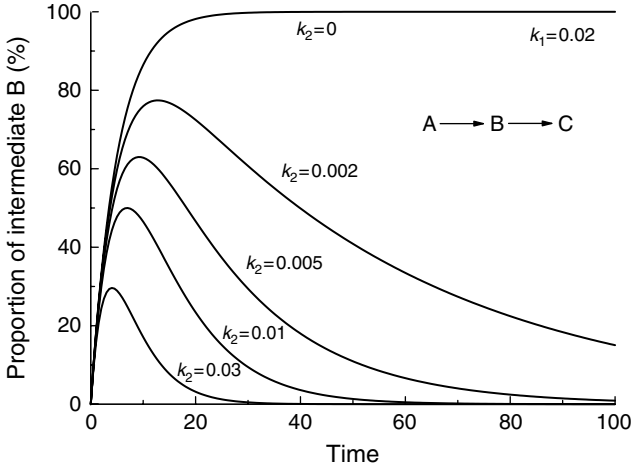


**Figure 1.** Simulation of two sequential first-order reactions. A simple reaction scheme A → B → C can be solved analytically using the equation in the text. This set of progress curves defines the amount of B (expressed as a percentage of the starting material), for a fixed value of $k_{A\rightarrow B}$ and different values of $k_{B\rightarrow C}$. The rate constants and *x*-axis are expressed in undefined units of time to indicate the generality of the analysis.

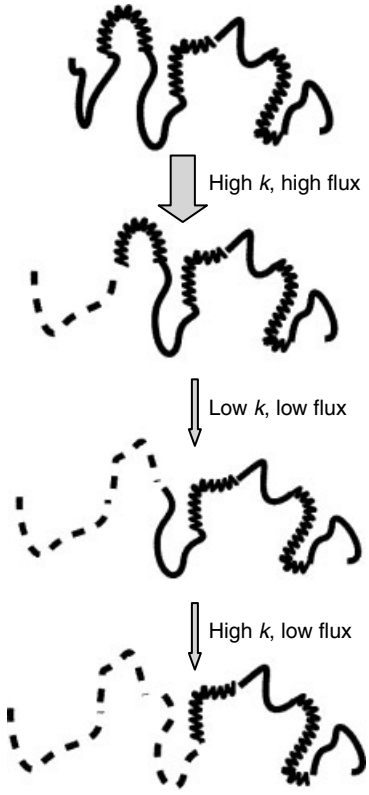


**Figure 2.** Exoproteolysis as a conformational probe. The model proposed by Villanueva *et al.*[7] assumes that proteins comprise alternating regions of sequence that are either susceptible to or resistant to exoproteolytic attack. However, a distinction must be made between the susceptibility to proteolysis (defined as the first-order rate constant, *k*) and the flux through the pathway to generate each intermediate. Although late stages can be characterized as having an inherently high susceptibility, flux through that process is largely influenced by the availability of substrate, a function of prior proteolytic steps.

JMS

full of water, and has a large hole through which water flows to the second pail—this large hole represents a high rate of proteolysis. The second pail has a small hole (a low rate of proteolysis). It will be intuitively obvious that the second pail (a stable degradation intermediate) will now fill with water, and that water will in turn leak slowly into the third pail. If the third pail has a large hole (it is proteolytically susceptible), it is again intuitive that the third pail is never going to become full—the pool of this particular unstable intermediate will be small. Finally, a fourth pail, representing a resistant intermediate, would need a very small hole in order that it will accumulate significantly, given the low rate of entry of material from the slow 'upstream' processes.

Analytical solutions to series of sequential first-order reactions become progressively more difficult to derive as the number of reactions increases. A valid and flexible alternative is to model such reactions in a kinetic simulation package, such as SAAM II (http://www.saam.com). In such a system, any number of linear or branched first-order pathways can be simulated by construction of a graphical metaphor (Fig. 3). The values of the successive rate constants can be fixed, the entire system solved and the amounts of each successive intermediate recovered as a function of reaction time. In the example shown (Figure 3) there are nine pools. The first (pool 1) is the starting material, and is assumed to be proteolytically susceptible. Pools 2, 4, 6 and 8 are all assumed to be relatively resistant to proteolysis (flanked by 'stop sites'), whereas pools 3, 5, 7 and 9 are assumed to be very susceptible to further digestion. The odd-numbered rate constants are high, and the even-numbered are assumed to be less than the odd-numbered values, consistent with the proteolytic resistance of the even numbered pools.

Using this model, multiple scenarios can been explored, and two are described in detail here. In model 1, alternating reactions have high and low rate constants. The model makes all fast rate constants ($k_f$), representing digestion of proteolytically susceptible species, adopt the same value. Similarly, all slow rate constants ($k_s$, reflecting digestion of the proteolytically resistant species) are set to the same value. The proteolytic sequence is the simplest, comprising

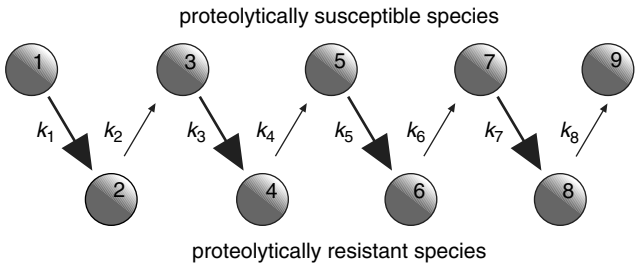


**Figure 3.** A complex model of sequential proteolysis. A nine-pool model of proteolysis, with first-order mass transfer between each pool, was defined in the SAAM II package. Pool 1 is the starting material, and is removed to pool 2 with a high rate constant ($k_1$). In turn, pool 2 is proteolysed to pool 3 by a slow process (rate constant = $k_2$). Pools with even numbers reflect proteolytically resistant species and those with odd numbers represent proteolytically susceptible intermediates.

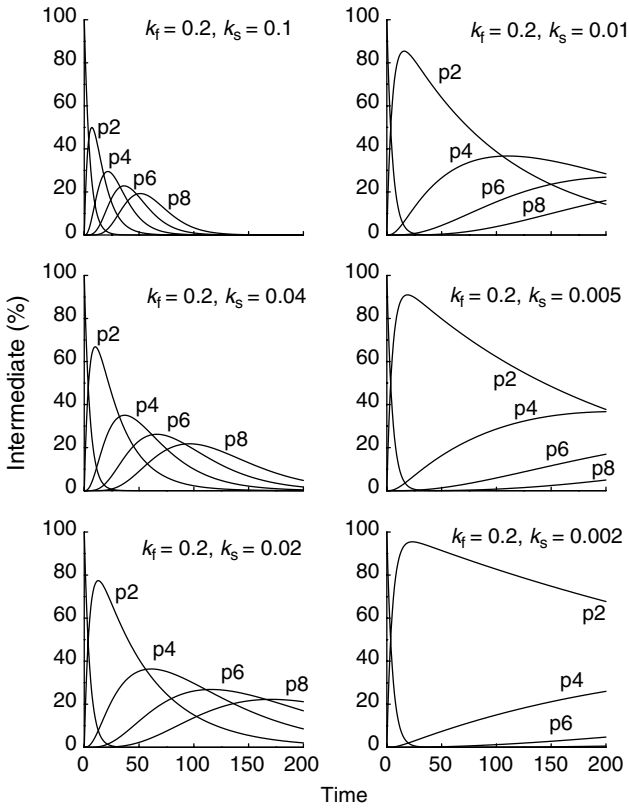


**Figure 4.** Simulation of the nine-pool scheme: I. The nine-pool model in Fig. 3 was simulated in the SAAM II package for different values of rate constants for the fast and slow steps. In this simulation, the rate constant for the fast steps ($k_f$) were set to be equal at $0.2t^{-1}$. The values for the slow rate constants ($k_s$) were again set to be the same for each step, but were allowed to adopt different values. The loss of starting material (pool 1) and the four stable intermediate pools were plotted for the duration of the simulation. Pool 1 is not labeled on every trace, as it always adopts the same trajectory.

successive alternating fast and slow reactions. The pool sizes of the proteolytically resistant, even-numbered pools are plotted in Fig. 4 (the pools sizes of the proteolytically susceptible species are so small that they are not displayed). For this particular simulation, the value of $k_f$ is maintained at $0.2t^{-1}$ and the value of $k_s$ is varied between $0.1t^{-1}$ (half of $k_f$) to $0.002t^{-1}$ (1% of the fast rate constant). When the fast and slow rate constants are very similar, there is a transient appearance of each stable pool in turn, but the amounts of material in each pool diminish at each successive stage in the reaction. However, this situation, where the rate constants are similar for fast and slow processes, does not reflect a structure-based influence on proteolytic rate—the stable and susceptible forms are equally well digested. This behaviour is in direct contradiction to that proposed.[7] As the value of $k_s$ becomes a smaller proportion of $k_f$, the behaviour changes, and it becomes increasingly difficult to accumulate substantial amounts of any intermediate other than the first one to be generated. This is because the rate of formation of the first stable product is high, but its rate of removal is very low. Hence, under conditions in which there is a strong influence of structure on rate of proteolysis, the later 'stable'

intermediates can never accumulate. The situation becomes even more exaggerated if the rate constants for the slow processes vary (Fig. 5). If, for example, $k_2$ is set to 10% of the value of all successive slow rate constants, pool 2 becomes a major reservoir of mass, and no subsequent intermediates can accumulate to any significant extent.

It might be argued that it would be feasible to invoke a scheme where the values of $k_f$ and $k_s$ were significantly different, in accordance with the model originally proposed, but that each successive value of $k_s$ was smaller than the one preceding it, creating a series of bottlenecks whereby each stable pool could accumulate (Figure 6). However, even this assumption creates difficulties. First, there are very few time points on the progress curve where it would be possible to recover significant quantities of all stable intermediates. Second, a proteolytic model in which each product becomes progressively more resistant to digestion is inherently unlikely, as the gradual loss of higher order structure would usually be considered to render the products more likely to be digested.

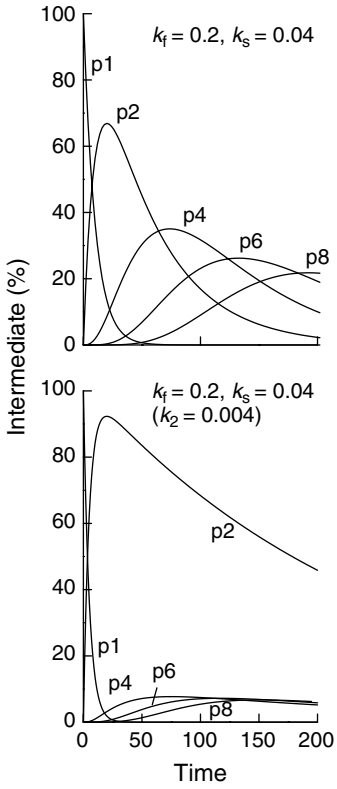


**Figure 5.** Simulation of the nine-pool scheme: II. The nine-pool model in Fig. 3 was simulated in the SAAM II package for different values of rate constants for the fast and slow steps. In this simulation, the rate constant for the fast steps ($k_f$) were set to be equal at $0.2t^{-1}$. The values for the slow rate constants ($k_s$) were set to be the same for each step at $0.04t^{-1}$ (top panel). In the bottom panel, the first slow step was set to a lower value of $0.004t^{-1}$. The loss of starting material (pool 1) and the four stable intermediate pools were plotted for the duration of the simulation.

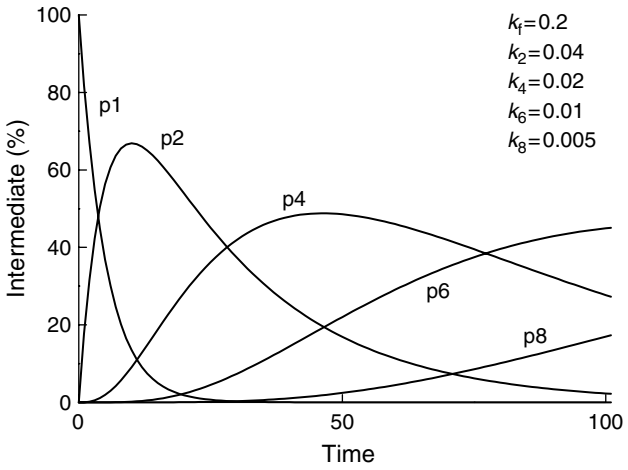


**Figure 6.** Simulation of the nine-pool scheme: III. The nine pool model in Fig. 3 was simulated in the SAAM II package for different values of rate constants for the fast and slow steps. In this simulation, the rate constant for the fast steps ($k_f$) were set to be equal at $0.2t^{-1}$. The values for the slow rate constants ($k_s$) were set to be progressively lower at $0.04t^{-1}$, $0.02t^{-1}$, $0.01t^{-1}$ and $0.005t^{-1}$. The loss of starting material (pool 1) and the four stable intermediate pools were plotted for the duration of the simulation.

## CONCLUSIONS

Whilst exoproteases may have some utility as structural probes in specific instances, there is a need for caution in generalizing the model. Ideally, sufficient time samples of a reaction mixture would be taken to allow for a detailed reconstruction of the course of the digestion, trapping each successive stable intermediate at an appropriate time. Then, a package such as SAAM II can be used to fit a proteolytic model to the data set, from which the values of the different rate constants can be extracted, as a rigorous test of the model. Of course, this in turn demands that the mass spectrometric analysis of each intermediate be conducted in such a way that the intermediates can be carefully quantified, a situation that is difficult to attain either with MALDI or electrospray sources. Indeed, it might be argued that densitometric analysis of one-dimensional with sodium dodecyl sulfate gels of the digestion mixture would yield more reliable data.

Finally, there remains the problem of explaining the data observed previously.[7] It can be suggested that the target proteins might have undergone limited endoproteolytic digestion at susceptible sites, followed by exoproteolytic trimming of the products. For this to be the case, the preparations of exoproteases would have been contaminated with trace amounts of endoproteases. Although this is conjectural, it is not without precedent[8] and is readily tested by, for example, zymography.[9]

### Acknowledgement

The author's research is supported by the Biotechnology and Biological Sciences Research Council.

JMS

## REFERENCES


1. Hubbard SJ, Beynon RJ. Proteolysis of native proteins as a structural probe. In *Proteolytic Enzymes. A Practical Approach*, Beynon RJ, Bond JS (eds). Oxford University Press: Oxford, 2001; 233.
2. Hubbard SJ. The structural aspects of limited proteolysis of native proteins. *Biochim. Biophys. Acta* 1998; **1382**: 191.
3. Fontana A, Polverino de Laureto P, De Filippis V, Scaramella E, Zambonin M. Probing the partly folded states of proteins by limited proteolysis. *Fold. Des.* 1997; **2**: R17.
4. Ellison D, Hinton J, Hubbard SJ, Beynon RJ. Limited proteolysis of native proteins: the interaction between avidin and proteinase K. *Protein Sci.* 1995; **4**: 1337.
5. Wu C, Robertson DHL, Hubbard SJ, Gaskell SJ, Beynon RJ. Proteolysis of native proteins. Trapping of a reaction intermediate. *J. Biol. Chem.* 1999; **274**: 1108.
6. Hubbard SJ, Beynon R, Thornton JM. Assessment of conformational parameters as predictors of limited proteolytic sites in native protein structures. *Protein Eng.* 1998; **11**: 349.
7. Villanueva J, Villegas V, Querol E, Aviles FX, Serrano L. Protein secondary structure and stability determined by combining exoproteolysis and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *J. Mass Spectrom.* 2002; **37**: 97.
8. Place GA, Beynon RJ. The chymotrypsin-catalysed activation of bovine liver glutamate dehydrogenase. *Biochem. J.* 1982; **205**: 75.
9. Beynon RJ, Oliver S. Avoidance of proteolysis in extracts. *Methods Mol. Biol.* 1996; **59**: 81.