Insulin: HPLC Mapping of Protease Digestion Products

A Biochemistry Laboratory Experiment

Martha M. Vestling
Department of Chemistry and Biochemistry, University of Maryland Baltimore County, Baltimore, MD 21228

Proteases are to proteins what restriction enzymes are to nucleic acids. Both proteases and restriction enzymes catalyze the hydrolysis of specific bonds in large molecules producing an array of smaller pieces. Proteases catalyze the hydrolysis of certain amide bonds in proteins while restriction enzymes catalyze the hydrolysis of certain phosphate bonds in nucleic acids. One way to characterize a macromolecule is to describe the fragments that result from a given reaction. By separating and counting the fragments, a pattern emerges. The pattern is called a map and the process, mapping.

For proteins, the fragments are peptides, and the most convenient technique for separating and counting them is high performance liquid chromatography (HPLC). The standard protocol calls for a reverse-phase column, gradient elution, and ultraviolet detection. This methodology has become increasingly important as the market for biotechnology products grows. Proteins, particularly those targeted for human use, require extensive purity checks (*1*, *2*).

The HPLC and protease experiment in this report, which will be described in detail below, is meant to introduce biochemistry students in one laboratory period to an important protein characterization protocol that can be used with other techniques for sequencing but is used regularly by itself to compare one batch of a protein with another batch of the same protein. In this experiment the protocol is applied specifically to the mapping of bovine insulin. The advantage of using insulin or other small polypeptides in testing protease digestions and HPLC procedures is that the resulting peptide mixtures are relatively simple and the HPLC data easily understood. The progress of a digestion can be monitored by watching the signal from the starting polypeptide disappear. And the expected number of peaks can be predicted from the original protein. The separation of simple peptide mixtures takes less HPLC time than separation of complex mixtures, so experiments can be worked into an undergraduate laboratory period or used routinely to test the viability of the protease.

Complex peptide mixtures require additional analysis to be understood even after HPLC separation. Generally protein laboratories either subject the individual peptides to Edman degradation (*3*) or to amino acid analysis, or subject partially separated mixtures of peptides to mass spectrometry (*4*) in order to identify the peptides. For simple mixtures from a known polypeptide, N-group determination may be all that is needed to completely identify each peptide.

The structure of bovine insulin (*5*) is shown in Figure 1. Five standard proteases and their reaction conditions are listed in Table 1. All five are commonly used in protein analyses. The hydrolysis reaction the proteases catalyze is shown in Figure 2. For bovine insulin, trypsin and *S. aureus* V8 protease are particularly useful while endoproteinase Asp-N is not and pepsin and α-chymotrypsin produce more complex mixtures than is desirable. Trypsin's specificity for catalyzing hydrolysis of amide bonds whose carbonyls come from arginine (R) and lysine (K) means only two amide bonds in insulin are expected to react leading to a very simple peptide mixture. *S. aureus* V8 protease (also called

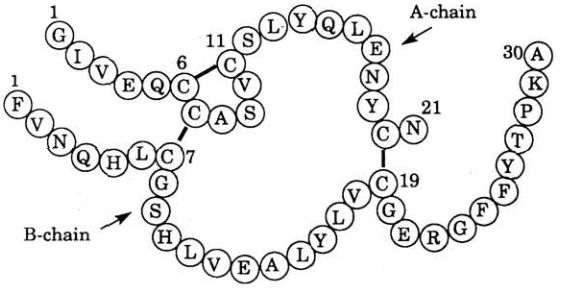

Figure 1. The primary structure of bovine insulin (5). The solid bars indicate disulfide bonds between the cysteine residues.

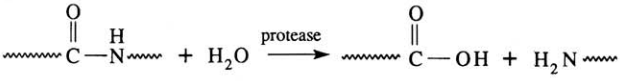

Figure 2. The hydrolysis reaction catalyzed by proteases.

Table 1. Protease Reaction Conditions

Protease	Specificity[a]	Solvent[b]	Reaction Buffer[c]	Conditions	Ref.
trypsin	R, K (C side)	0.1 mM $CaCl_2$	0.1 M NH_4HCO_3 0.1 mM $CaCl_2$	37 °C, 2–4 h	*7*
S. aureus V8 protease	E (C-side) E + D in other buffers	water	0.1 M NH_4HCO_3	37 °C, 4–6 h	*8*
endoproteinase Asp-N	D (N-side)	water	0.1 M NH_4HCO_3	37 °C, 4 h	*9*
pepsin	C-side of F, Y, W, L, M, H, Q	water	0.1% TFA	25 °C, 2 h	*10*
α-chymotrypsin	C-side of F, Y, W, L, M, H, Q	0.1 mM $CaCl_2$	0.1 M NH_4HCO_3	37°C, 2–4 h	*7*

[a]The one letter amino acid abbreviations are used. C-side refers to the C-terminal side of the amino acid residue indicated while N-side refers to the N-terminal side of the amino acid residue indicated.

[b]Dissolve each protease in its solvent (1.0 mg/mL) and ice immediately.

[c]Insulin should be dissolved in the reaction buffer to make a 1.0 mg/mL solution.

endoproteinase Glu-C) reacts in ammonium bicarbonate only with amides in which the carbonyl has been contributed by glutamic acid (E). Since insulin contains only four E's, again the digest mixture should contain only a few peptides. Endoproteinase Asp-N requires an aspartic acid (D) residue to be present in the amino part of an amide. Since insulin does not contain any D's, its hydrolysis will not be catalyzed by endoproteinase Asp-N. Both α-chymotrypsin and pepsin pick aromatic and hydrophobic amide bonds to hydrolyze, so cuts after F, Y, W, L, M, H, and Q can be expected. For insulin, that means lots of little pieces and a complex HPLC pattern.

In general, protease digests do not disrupt disulfide bonds, so insulin digested by *S. aureus* V8 protease should lead to four peptides while digestion of reduced insulin should lead to six peptides. Therefore, insulin, native and reduced, in the presence of five proteases can be used to generate up to 10 different HPLC traces. For larger laboratory classes or for different experiments in different years, the digestion and HPLC analysis of relatively small and definitely cheap proteins such as calcitonin (bovine), cytochrome c (equine heart), and ribonuclease A (bovine) can be substituted for insulin.

Insulin was the first protein to be sequenced (*6*), and the work, which earned F. Sanger his first Nobel prize in 1958, involved partial acid hydrolysis, protease (pepsin) catalyzed hydrolysis, chromatographic separation of peptides, and many N-group determinations, took several years to complete. Today with the wider variety of proteases, HPLC, and automatic sequencers and mass spectrometers, the sequence of insulin can be easily checked in a couple of days or less, making the following student experiment feasible.

Procedure

Chemicals

The suppliers of the necessary chemicals are shown in square brackets: ammonium bicarbonate, calcium chloride, acetonitrile (HPLC grade) [J. T. Baker], trypsin (bovine pancreas, TPCK treated), α-chymotrypsin (bovine pancreas, TLCK treated), pepsin (porcine), *S. aureus* V8 protease, dithiothreitol [Sigma], trifluoroacetic acid [Aldrich].

Protease Digestion

The experimental protocol is summarized in Table 1. The enzyme solutions (1.0 mg/mL) are prepared first. We weigh out 1–2 mg of protein and then add the corresponding 1–2 mL of solvent using an automatic delivery pipet so that the final solution is 1.0 mg/mL. The contents are mixed by spinning the tube or vial in an ice bath. The insulin solutions are prepared in the same manner using the reaction buffers indicated in Table 1. Each reaction is started by adding 20 μL of enzyme solution to an insulin solution which is then capped and placed in a 37 °C bath. After incubation for 4 h, the reaction is placed on ice, ready for HPLC analysis. The digests can be stored at −70 °C for several months with no apparent degradation. Treatment of 100-μL aliquots of the digests with 10 μL of 10 mg/mL dithiothreitol for 5 min will reduce the disulfide bonds.

HPLC Parameters

This experiment requires an HPLC with gradient capability and works best with a variable ultraviolet wavelength detector. A reverse-phase C18 column (4.6 × 100 or 4.6 × 250 mm) (60–100 Å pores); detector at 215 nm (254 nm can be used if the sample size is increased 20–50 times); flow rate = 1.0 mL/min; mobile phase A = 0.1% TFA in water; mobile phase B = 0.1% TFA in acetonitrile. (HPLC grade water and acetonitrile should be used. HPLC water can be prepared from good deionized water with a 0.45 μm nylon filter.) Gradient = 10–60% B over 20 min; 20 μL injections with adjustment of the range on the detector to fit the sample. (The reason the chromatograph and the C18 column's manufacturers are not specified here is that chromatographs and columns from several manufacturers have been tried and found to work.)

Results and Discussion

Figure 3 contains the reverse-phase HPLC maps for unreacted insulin, insulin incubated with trypsin, and insulin incubated with *S. aureus* V8 protease. The patterns are quite different. Since reverse-phase chromatography separates peptides approximately by size and the intensity

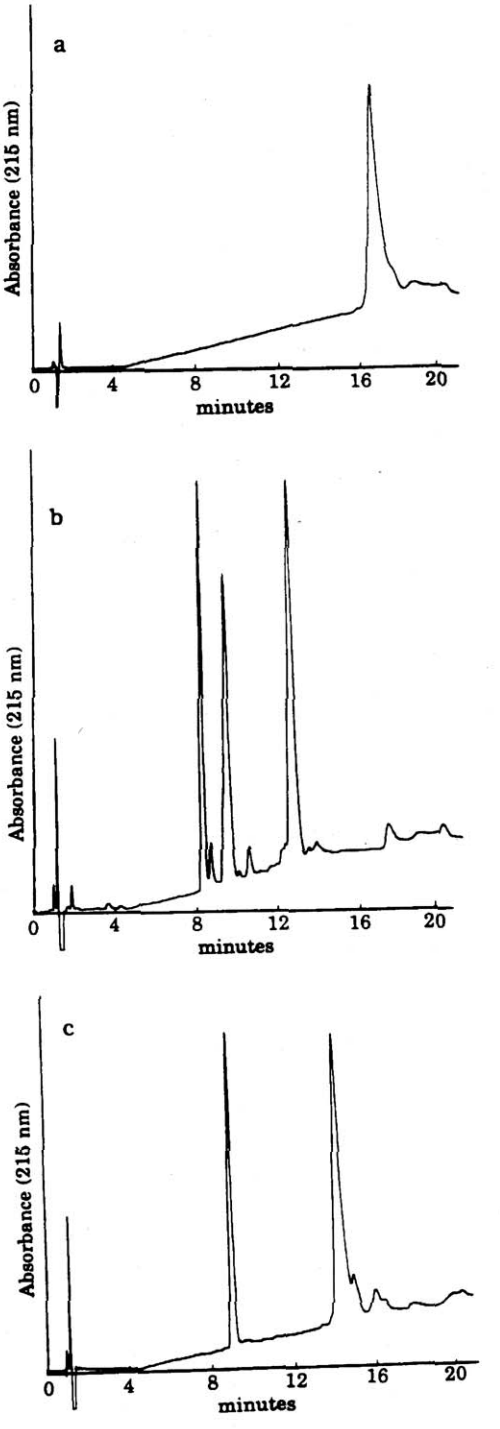

Figure 3. Chromatographic tracers forc(a)insulin, (b) insulin/S. aureus V8 protease reaction mixture, and (c) insulin/trypsin mixture. See text for details

of the absorption of 215 nm light is approximately proportional to the number of amide bonds, peaks from different-sized peptides should be separated and should have different heights as seen in Figure 3. All amide bonds are not hydrolyzed at the same rate, and some partially hydrolyzed products are generally seen. Thus equimolar amounts of each peptide are not expected, contributing to further variations in peak height, and minor peaks representing partially hydrolyzed products are observed. However, HPLC maps may be reproducibly obtained when all the experimental variables are held constant.

The 10–60% linear gradient is not necessarily the best gradient for a particular digest, but, in general, it is a good starting place for analyzing peptides as it immediately gives the operator an estimate of how complex the mixture is. For insulin digests, the linear gradient works fine.

As long as insulin gives a reproducible peak and as long as a digest gives peaks when injected into the HPLC, the data is suitable for student analysis. They are asked to predict the number of peaks, compare their results with the prediction, and to suggest ways to test their conclusion. We have also prepared the enzyme solution for them and then asked them to determine which of the enzymes in the table it was. The idea is to get them to look at their data and at the structures.

In the teaching laboratory, despite common directions, a variety of traces were acquired. It is clear that extraneous proteins and peptides and other chemicals that absorb at 215 nm are easily added to the reactions unless care is taken. The use of clean equipment is imperative, and proper balance techniques are critical. Keeping hands off pipet tips, the working blades of spatulas, and interestingly, off the plungers of the blunt-tipped syringes used for HPLC requires some coaching. Students with microscale organic laboratory experience should have some advantages here.

Students are asked to obtain a duplicate HPLC trace to be sure that what they record reflects the digest, not the condition of the HPLC. If only one HPLC is available for student use, students can be scheduled so that two or three per laboratory period can analyze their digests while the rest of the class works on other experiments. No one expects students to master HPLC in part of one laboratory period, but some introduction is becoming increasingly important.

Because this experiment requires a gradient HPLC, most teaching laboratories will have to plan around two to three students per HPLC per laboratory period with the digests being stored at –70 °C until needed. Intructors should plan on about one hour to start up the HPLC's and run baseline gradients before each laboratory period. Preparation of the HPLC mobile phases and the protease reaction buffers requires time also but can be done once each semester and does not need to be done again before every laboratory period. Refrigeration of these solutions between laboratory periods is probably a good idea.

Each of the peaks in Figure 3b and 3c was collected by the author and characterized by molecular weight, as determined by mass spectrometry. The molecular weights are listed in Table 2 in the order in which they eluted from the reverse-phase column. The peptides were identified by comparing the molecular weights with the those predicted based on the known sequence of insulin and the known specificity of the protease. The three main peaks in Figure 3b were easily identifiable, as were the two main peaks in Figure 3c. The expected *S. aureus* V8 protease peptide [A1–A4] with a molecular weight of 416 was found in the 3–4 min HPLC fraction. The shoulder on the second trypsin peak was identified as an incomplete hydrolysis product. Other aberrations in the baselines were not identified.

Table 2. Characterization of Two Protease Reaction Mixtures by Mass Spectrometry

Protease	molecular weight [a,b]	peptide assignment [c]
S. aureus V8 protease	416	[A1–A4]
	1376	[A18–A21]+[B14–B21]
	1086	[B22–B30]
	2923	[A5–A17]+[B1–B13]
trypsin	858	[B23–B29]
	4822	[A1–A21]+[B1–B22]
	5662 [d]	[A1–A21]+[B1–B29]

[a] As determined by fast atom bombardment mass spectrometry (*4*).
[b] Peptides are listed in the order they eluted from the reverse phase column.
[c] "+" indicates two peptides joined by a disulfide bond.
[d] Shoulder on previous peak.

The protease and HPLC techniques described above for insulin are useful in characterizing larger proteins. They are, after all, just two specific tools in the standard chemistry strategy which calls for breaking large molecules into recognizable smaller ones.

Acknowledgment

The author acknowledges the encouragement of Catherine Fenselau in whose laboratory this biochemistry experiment was developed. Financial support was provided by SUNY College at Brockport in the form of a sabbatical and by a Research Opportunity Award from the National Science Foundation.

Literature Cited

1. Chloupek, R. C.; Harris, R. J.; Leonard, C. K.; Keck, R. G.; Keyt, B. A.; Spellman, M. W.; Jones, A. J. S.; Hancock, W. S. *J. Chromatog.* **1989**, *463*, 375–396.
2. Furuya, M.; Akashi, S.; Hirayama, K. *Biochem. Biophys. Res. Commun.* **1989**, *163*, 1100–1106.
3. Yamashita, H.; Theerasilp, S.; Aiuchi, T.; Nakaya, K.; Nakamura, Y.; Kurihara, Y. *J. Biol Chem.* **1990**, *265*, 15770–15775.
4. Biemann, K. *Biomed. Environ. Mass Spectrom.* **1988**, *16*, 99–111.
5. Dayhoff, M. O. *Atlas of Protein Sequence and Structure*; National Biomedical Research Foundation: Washington D. C., 1972; Vol. 5, p D-209.
6. Sanger, F. *Science* **1959**, *129*, 1340–1344.
7. Titani, K.; Sasagawa, T.; Resing, K.; Walsh, K. A. *High Performance Liquid Chromatography of Proteins and Peptides*; Academic Press: NewYork, 1982; pp 23–27.
8. Rose, K.; Savoy, L.; Simona, M. G.; Offord, R. E.; Wingfield, P. *Biochem. J.* **1988**, *250*, 253–259.
9. Geuss, U.; Schaffer, M.; Tschakert, J.; Fischer, S.; Kresse, G. *J. Protein Chem.* **1990**, *9*, 299–300.
10. a) Morris, H. R.; Pucci, P. *Biochem. Biophys. Res. Commun.* **1985**, *126*, 1122–1128; (b) Toren, P.; Smith, P.; Chance, R.; Hoffman, J. *Anal. Biochem.* **1988**, *169*, 287–299; (c) Sun, Y.; Smith, D. L. *Anal. Biochem.* **1988**, *172*, 130–138.