Laboratory for Computational Proteomics

www.FenyoLab.org

E-mail: Info@FenyoLab.org

Facebook: *NYUMC Computational Proteomics Laboratory*

Twitter: *@CompProteomics*

CABIOS APPLICATIONS NOTE

Vol. 13 no. 6 1997
Pages 617-618

A software tool for the analysis of mass spectrometric disulfide mapping experiments

David Fenyö

The Rockefeller University, New York, NY 10021, USA

Received on April 2, 1997; revised and accepted on July 25, 1997

Disulfide cross-linkages stabilize the structure of many proteins (Creighton, 1993). Therefore, knowledge of the disulfide linkages in a protein provides useful, albeit partial, high-resolution structural information which can be obtained from experiments that are much less time consuming than the X-ray crystallography or NMR spectroscopy experiments necessary for complete structure elucidation. Determination of the disulfide bonds in recombinant or synthetic proteins is also important, since formation of the correct disulfide cross-linkages is an indication of proper folding and function. In addition, protein folding can be studied by determining the disulfide bonds in folding intermediates (Creighton, 1993).

The disulfide linkages in a protein can be determined by proteolytically digesting the protein under conditions where the disulfide bonds are stable, separating the proteolytic peptides, and identifying those that contain disulfide linkages. The identification can be performed by amino acid analysis, amino-terminal sequencing, or mass spectrometry. Sometimes, a combination of these techniques is used. The advantages of using mass spectrometry for disulfide mapping (Smith and Zhou, 1990) include its speed, sensitivity, specificity and ability to analyze mixtures.

In a typical experiment for determining the disulfide linkage pattern, the protein is digested with a proteolytic enzyme and the reaction mixture analyzed with mass spectrometry before and after reduction of the disulfide bonds. The two mass spectra are compared and the masses of the peptides that contain disulfide bonds are determined. If the mass spectra are complicated, the reduced material can be alkylated and a third mass spectrum acquired to identify the cysteine-containing peptides. Also, the non-reduced sample can be alkylated to determine the free cysteines in the protein. In cases where the protein of interest is small, contains few disulfide linkages, and can be satisfactorily digested with a protease having high specificity, it is frequently feasible to identify the peptides that are cross-linked simply from the mass(es) of the cross-linked peptides. In cases where more information is necessary for unambiguous identification of the interconnected peptides, components of the proteolytic digestion mixture can be separated and the fractions analyzed by mass spectrometry before and after reduction of the disulfide bonds. Alternatively, the disulfide-containing peptides can be isolated in the mass spectrometer and caused to fragment. One difficulty with this method is that proteins frequently have tightly folded domains that resist proteolysis by enzymes with high specificity. This problem can be solved by using enzymes with low specificity (e.g. thermolysin) or enzymes that have both low specificity and the capability of functioning under normally denaturing conditions (e.g. pepsin). However, the use of enzymes with low specificity requires additional constraints for the identification of the connected proteolytic peptides. Such constraints can be provided, for example, by amino-terminal sequencing (Burman *et al.*, 1989) or mass spectrometric fragmentation (Bean and Carr, 1992). Another potential limitation of the proteolytic strategy for disulfide mapping is that enzymes do not always cut between all the half-cystine residues present in the protein, and so yield complex peptides that contain more than two peptides connected by multiple disulfide bonds. Some of

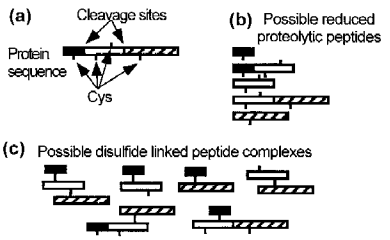

Fig. 1. (**a**) A protein sequence containing two enzymatic cleavage sites and four cysteine residues that may or may not be involved in disulfide bonds. (**b**) Enzymatic cleavage under reducing conditions results in five possible peptide fragments. (**c**) Calculation of cleavages (allowed by the cleavage rules of the enzyme of interest) under conditions that leave the disulfide bonds intact results in six possible disulfide-linked peptides. The peptides containing two free cysteines can form intra-chain disulfide bridges (not shown).

E-mail: fenyod@rockvax.rockefeller.edu

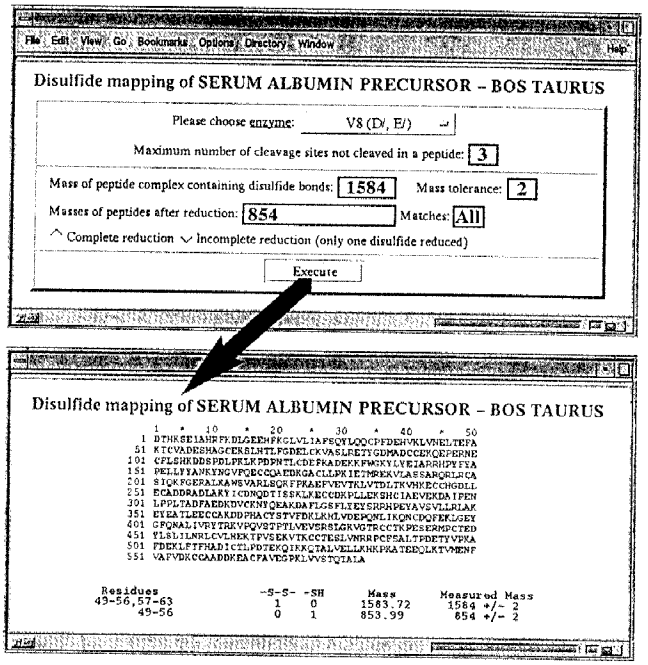

Fig. 2. This example shows a search performed with the mass of a disulfide-linked multi-chain peptide (1584 ± 2 Da) from a hypothetical V8 protease (cleaves at the C-terminal side of glutamic and aspartic acid residues) digest of bovine serum albumin and the mass of one of its component peptides (854 ± 2 Da).

the resulting ambiguities in disulfide bond linkage for complex proteolytic peptides containing more than two interconnected peptides can be resolved by strategies that incorporate partial reduction of the disulfide bonds (Gray, 1993).

Although mass spectrometric disulfide mapping is potentially fast, the experimental results are usually analyzed manually—a time-consuming task even for small proteins and prohibitively so for larger proteins. Computer programs have been written to simplify the analysis (e.g. Smith *et al.*, 1990; Beavis, 1996), but these only provide partial solutions to the disulfide mapping problem. Here, we present a general software tool for the complete analysis of mass spectrometric disulfide mapping experiments. The algorithm allows for rapid disulfide linkage mapping of proteins that have undergone complete or partial digestion using enzymes with either high or low specificity. In addition, the algorithm allows for the incorporation of data from mass spectrometric fragmentation experiments.

Initially, a list is made from the amino acid sequence (Figure 1a) of all possible linear reduced proteolytic digestion products that contain cysteine residues and can be generated from treatment of the protein with the enzyme used, i.e. peptides allowed by the specificity of the enzyme (Figure 1b). All possible disulfide-linked multi-chain peptides (containing any number of proteolytic peptides) are identified and their masses are calculated (Figure 1c). The calculated masses are compared with the experimentally determined masses and matches are found, allowing the amino acid sequences of the disulfide-linked multi-chain peptides to be deduced. Finally, the algorithm tests whether the disulfide-linked multi-chain peptides contain component peptide chains with masses corresponding to the measured masses of the reduced peptides.

The disulfide mapping tool is part of the PROWL environment (Fenyö *et al.*, 1996). To reach the disulfide mapping tool, first select ProteinInfo and search for the protein of interest in a selected protein sequence database, and then click on 'sequence'. Select Disulfide mapping under Choose task. The search constraints can be the mass of the disulfide-linked multi-chain peptide either with or without the masses of one or more of the component peptide chains (Figure 2). The proteolytic enzyme can be selected from a list or the specificity of the digestion can be specified, allowing for the use of multiple enzymes in the search. The degree of completeness of the digestion can be selected, making the searches fast for cases where complete digestion can be achieved, while permitting the analysis of incomplete digestion products when necessary.

Acknowledgements

This work was supported by the National Science Foundation (grant 9630936) and the National Institute of Health (grant RR00862). The authors acknowledge many fruitful discussions with Brian T.Chait, Ron C.Beavis, Júlio C.Padovan, Jun Qin and Salvatore Sechi.

References

Bean,M.F. and Carr,S.A. (1992) Characterization of disulfide bond position in proteins and sequence analysis of cystine-bridged peptides by tandem mass spectrometry. *Anal. Biochem.*, **201**, 216–226.

Beavis,R.C. (1996) Protein Analysis Worksheet. Personal communication (URL http://www.proteometrics.com/software/paws.htm).

Burman,S., Wellner,D., Chait,B., Chaudhary,T. and Breslow,E. (1989) Complete assignment of neurophysin disulfides indicates pairing in two separate domains. *Proc. Natl Acad. Sci. USA*, **86**, 429–433.

Creighton,T.E. (1993) *Proteins. Structures and Molecular Properties*, 2nd edn. W.H.Freeman and Co., New York.

Fenyö,D., Zhang,W., Chait,B.T. and Beavis,R.C. (1996) Internet-based analytical chemistry resources: a model project. *Anal. Chem.*, **68**, 721–726.

Gray,W.R. (1993) Disulfide structures of highly bridged peptides: a new strategy for analysis. *Protein Sci.*, **2**, 1732–1748.

Smith,D.L. and Zhou,Z. (1990) Strategies for locating disulfide bonds in proteins. *Methods Enzymol.*, **193**, 374–389.