UNIVERSITÄT LEIPZIG

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Software Development for Determining Cross-Linking Sites in Proteins

Andrea Sinz[1] and Carsten Sinz[2]

[1]Biotechnological-Biomedical Center, Institute of Analytical Chemistry, University of Leipzig, Linnéstr.3, D-04103 Leipzig
[2]Symbolic Computation Group, WSI for Computer Science, University of Tübingen, Sand 13, D-72076 Tübingen

Introduction

Chemical cross-linking, a conceptually simple approach, has gained renewed interest in combination with mass spectrometric analysis of the reaction products by its possibilities to elucidate three-dimensional structure of proteins as well as interfaces among interacting proteins in their natural environment [1-3]. The identification of a large number of cross-linking sites from the complicated mixtures generated by chemical cross-linking however remains a daunting task. To facilitate determination and classification of cross-linking sites, we developed a novel software program named Calc-XL.

Program Description

Data Input

In Calc-XL, the user defines the protein sequence(s) to be cross-linked, digestion enzyme, maximum number of missed cleavage sites, mass shift caused by the applied cross-linker, and the amino acids, which are targeted by the cross-linking reagent. We included a feature that accounts for (partial or complete) modification reactions before the cross-linking reaction itself, e.g., introduction of sulfhydryl groups, which are necessary in some cases [3] in order to generate reactive sites and to modify the specificity of the cross-linking reaction.

Calculation of Cross-Linking Products

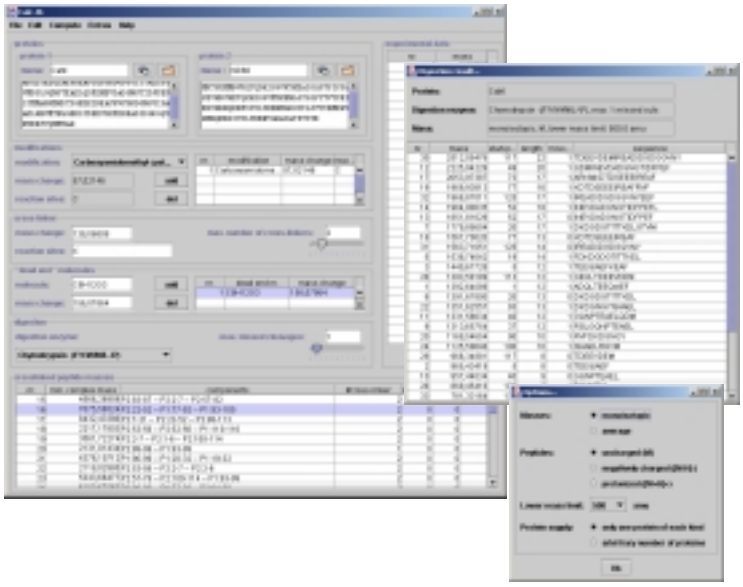
Calc-XL calculates molecular masses of all possible cross-linking products of the protein(s) (inter- and intramolecular products) with a given cross-linker to five decimal places. In contrast to already existing software [4], we are able to handle a more complex set of cross-linking products.
Calculation is performed by the Omega test, which combines new methods for eliminating equality constraints with an extension of Fourier-Motzkin variable elimination to integer linear programming [5].

In the first Omega test, the minimally possible complex mass $m_{c,min}$ (no additional cross-linkers, modifications, "dead end molecules") (cross-linking product A 1) is calculated for an experimental mass. At the same time the maximally possible complex mass $m_{c,max}$ (maximum number of additional cross-linkers, modifications, "dead end molecules") is calculated, and both complex masses are compared to experimental data. If no experimental mass falls into the $m_{c,min}$–$m_{c,max}$–window there is no match and another cross-linking product is created.

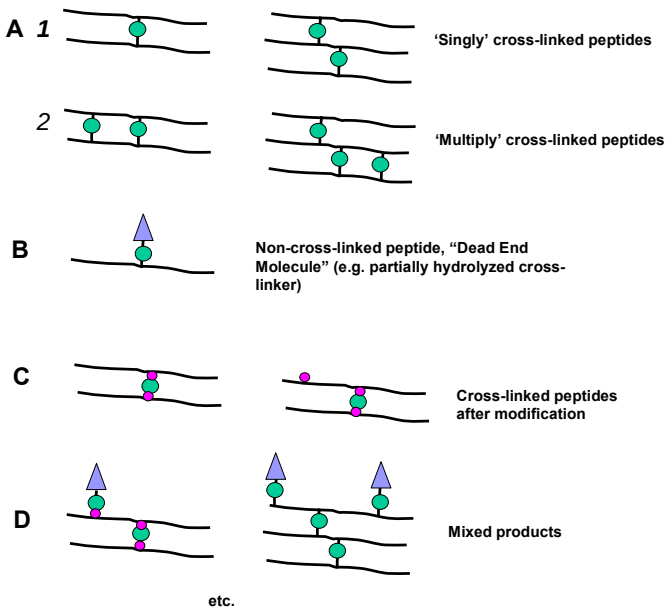
If there is a match of the cross-linking product with experimental data, a second Omega test is performed generating all possible cross-linking product configurations consisting of different "dead end molecules", additional cross-linkers, and varying modifications (cross-linking products A2, C, and D). If there is a mass match with experimental data within the given tolerance for mass accuracy, the match is written to the output list.

For correct assignment of the complex product mixtures created by chemical cross-linking, a small tolerance window for calculated and experimentally obtained mass data has to be chosen. In our case, we will be using data from Fourier Transform Ion Cyclotron Resonance mass spectrometry (FT-ICR MS), where an accuracy of better than 5 ppm is obtained routinely.

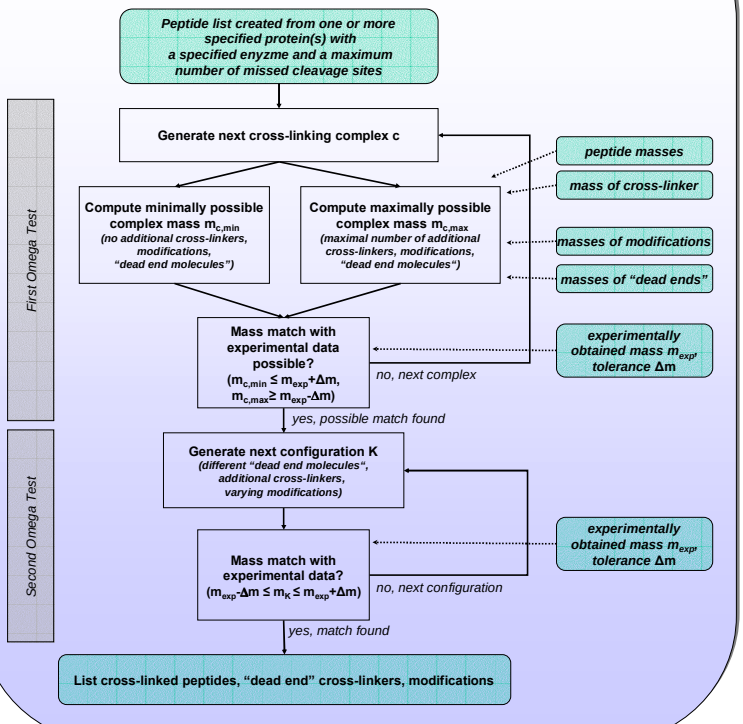
After calculation, Calc-XL classifies the results of possible cross-linking sites and gives a rating for the cross-linking site based on the sequence positions of cross-linkers in cross-linked peptides.



Cross-Linking Products



Program Scheme



References

[1] Young, M.M., Tang, N., Hempel, J., Oshiro, C.M., Taylor, E.W., Kunz, I.D., Gibson, B.W. and Dollinger, G. (2000) *Proc. Nat. Acad. Sci.* 97, 5802-5806.
[2] Bennett, K.L., Kussmann, M., Bjørk, P., Godzwon, M., Mikkelsen, M., Sørensen, P. and Roepstorff, P. (2000) *Protein Sci.* 9, 1503-1518.
[3] Sinz, A. and Wang, K. (2001) *Biochemistry* 40, 7903-7913.
[4] König, S., Zeller, M., Haberland, J., Gerke, V. and Højrup, P. (2001) in: *Proceedings of the 49th ASMS Conference on Mass Spectrometry and Allied Topics*, Chicago, IL.
[5] Pugh, W. (1992) *Communications of the ACM* 35 (8), 102-114.