

A Mixed-Integer Linear Optimization Framework for the Identification of Post-translational Modifications in Histone H3 using Electron Transfer Dissociation Tandem Mass Spectrometry

Peter A. DiMaggio*, Richard C. Baliban, and Christodoulos A. Floudas
Department of Chemical Engineering, Princeton University

Nicolas L. Young and Benjamin A. Garcia
Department of Molecular Biology, Princeton University

Session 3, Talk 2 (1:20 PM)

In recent years, there has been significant interest in the identification of histone post-translational modifications. It is well-known that histone proteins are key regulators of many important DNA processes in eukaryotes and recent studies have elucidated complex relationships between histone modifications and DNA damage response and repair. Early studies could only analyze these histone modifications on a site-by-site basis and as a result lose any connectivity information on the molecular level. The application of traditional reversed phase HPLC to separate the modified forms of histone H3 results in poor chromatographic resolution since all forms have very similar physical characteristics. Recent experimental advances [1] have resulted in the development of a novel high-throughput on-line liquid chromatography mass spectrometry method for the analysis of histone codes using a pH gradient and HILIC separation to elute the modified forms.

In this work, we present a novel methodology for the identification of the targeted post-translational modifications (PTMs) present in highly-modified proteins using mixed-integer linear optimization (MILP) and electron transfer dissociation (ETD). For a given ETD tandem mass spectrum, the rigorous set of modified forms that satisfy the mass of the precursor ion, within some tolerance error, are enumerated by solving a feasibility problem via mixed-integer linear optimization. The enumeration of the entire superset of modified forms enables the method to normalize the relative contributions of the individual modification sites. Given the entire set of modified forms, a superposition problem is then formulated using mixed-integer linear optimization to determine the relative fractions of the modified forms that are present in the multiplexed ETD or ECD tandem mass spectrum. The utility of the proposed method is demonstrated on a complex-mixture of highly-modified histone H3 for a single liquid chromatography mass spectrometry (LC-MS) experiment, where we show that the novel MILP framework is capable of consistently identifying primary and even secondary or lower level modified forms that would be missed by conventional identification methods.

[1] N.L. Young, P.A. DiMaggio, M.D. Plazas-Mayorca, R.C. Baliban, C.A. Floudas, and B.A. Garcia, in prep (2008).