

Computational Protein Structure Prediction from Mass Spectrometry Data

Steffen Lindert

Department of Chemistry and Biochemistry, Ohio State University, United States

Mass spectrometry-based methods such as covalent labeling, surface induced dissociation (SID) or ion mobility (IM) are increasingly used to obtain information about protein structure. However, in contrast to other high-resolution structure determination methods, this information is not sufficient to deduce all atom coordinates and can only inform on certain elements of structure, such as solvent exposure of individual residues, properties of protein-protein interfaces or protein shape. Computational methods are needed to predict high-resolution protein structures from the mass spectrometry (MS) data. Our group develops algorithms that use mass spectrometry data to elucidate protein structure. These algorithms can incorporate several different types of mass spectrometry data, such as covalent labeling, surface induced dissociation, and ion mobility. We developed scoring functions that assess the agreement of residue exposure with covalent labeling data, the agreement of protein-protein interface energies with SID data and the agreement of protein model shapes with collision cross section (CCS) IM measurements. Using our MS-guided algorithms, we have been able to correctly predict single subunit proteins as well as protein complexes from MS labeling, SID and IM data, in several cases outperforming the state-of-the-art structure prediction software AlphaFold.

References:

1. Drake, Z.C.; Seffernick, J.T.; Lindert, S. (2022), Protein complex prediction using Rosetta, AlphaFold, and mass spectrometry covalent labeling. *Nat Commun* 13 (1), 7846.
2. Seffernick, J.T.; Harvey, S.R.; Wysocki, V.H.; Lindert, S. (2019), Predicting Protein Complex Structure from Surface-Induced Dissociation Mass Spectrometry Data. *ACS Cent Sci* 5 (8), 1330-1341
3. Biehn, S.E.; Lindert, S. (2021), Accurate Protein Structure Prediction with Hydroxyl Radical Protein Footprinting Data. *Nat Commun* 12, 341