

精製タンパク質分子内に存在するジスルフィド結合パターンの検出と予測の最適化

黒田研究室

学籍番号: 17251002

茜 拓斗

【背景・目的】 液体クロマトグラフィー質量分析計(LCMS)は定性的及び定量的な分析に秀でた装置で、食品分野や環境分野など、幅広く使用されている。蛋白質科学分野においてはプロテアーゼ処理したペプチドを試料とし、その断片ペプチドを検出するために用いられる。

当研究室では精製タンパク質にタンパク質分解酵素を加えたものを試料とし、LCMS 測定を行うことで精製タンパク質が目的タンパク質と同位置にジスルフィド結合を形成しているかを確認している。目的タンパク質の同定の際、算出した m/z 値と測定値との照合を行うが、当研究室はこの工程が手作業であり極めて非効率的である。また特定のジスルフィド結合体に限定して同定しているため、存在し得るペプチドを見逃している可能性が高い。本研究では、タンパク質同定を自動化し全てのジスルフィド結合の組み合わせを試行することで、存在可能性の高いジスルフィド結合体を自動的に検出するシステムの構築を目的とする。

【手法】 測定器は、LCMS-8040(株式会社島津製作所)を使用した。ジスルフィド結合位置が既知である鶏卵白質リゾチーム(PDB:1DPX)を Pepsin 処理したものを試料とし、測定データを取得した。Blank 測定データ、試料を DTT 還元した測定データ(DTT データ)も取得し、ProteoWizard でファイルをプログラムで使用できる形式(.mzML)に変換した。次に、以下の処理を行うプログラムを作成した。

サンプル測定データと Blank 測定データを照合、一致する RT 値、m/z 値を持つピークを削除する(ピークデータ 1)。複数のアミノ酸配列(断片ペプチド)の全組み合わせを取得、各ペアの質量数、m/z 値を算出する。算出 m/z 値とピークデータ 1 を照合、一致する m/z 値を持つピークを抽出する(ピークデータ 2)。DTT データとピークデータ 2 を照合、一致する RT 値、m/z 値を持つピークを削除する(ピークデータ 3)。ピークデータ 3 から最頻値の RT を取得し、その RT 値を持つピークをグループ化していく。グループ内で配列ペア名、adduct の種類で分類する(Exist Table)。

最後に、Exist Table で出力されたピークの形状を目視で確認した。

【結果】 Exist Table 出力ピークの形状を目視で確認したところ、RT204 秒の seq1,seq2 結合体、seq3,seq4 結合体、RT206 秒の seq1,seq2 結合体、RT200 秒の seq1,seq2 結合体が存在する可能性が高いという結果が出た。試料として使用したリゾチームは seq1 と seq2、seq3 と seq4 でジスルフィド結合を形成していることより、適切なジスルフィド結合体のピークを検出できたと言える。

今後、結合体内にジスルフィド結合が2つ存在する場合の処理もアルゴリズムに追加する。また Blank データ、DTT データとの比較を更に詳細に行い、本システムの精度を高めていく。

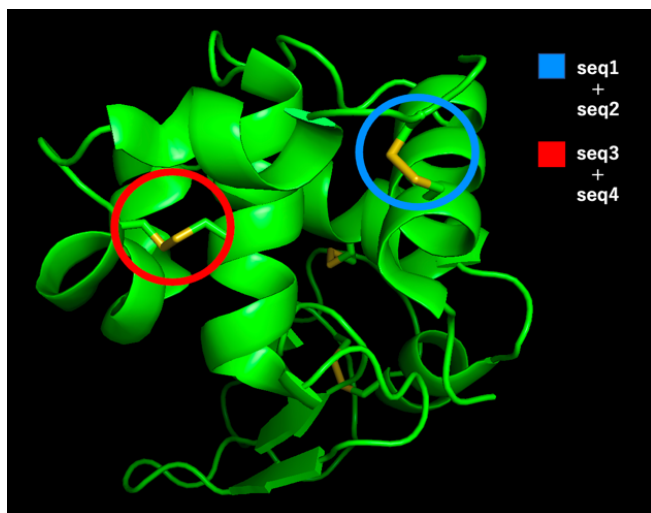


図1) 1DPX(PDB)ジスルフィド結合箇所

RT	name	adduct	mz_ave	Intensity_sum	count
204	seq1_seq2	(M+3H)3+	394.25	1074150	2
		(M+4H)4+	294.25	1022618	2
		(M+5H)5+	235.1	1064215	2
		(M+6H)6+	197.6	1056799	2
seq1_seq4		(M+10H)10+	163.15	1056071	2
		(M+6H)6+	275.4	516210	1
		(M+7H)7+	235.1	1064215	2
		(M+8H)8+	205.6	517280	1
		(M+9H)9+	163.15	1056071	2
seq3_seq4		(M+5H)5+	294.25	1022618	2
		(M+6H)6+	246.433333	1693441	3
		(M+7H)7+	209.5	630560	1
		(M+8H)8+	185.1	1715299	3
140	seq1_seq2	(M+10H)10+	119	6634013	10
		(M+9H)9+	130.15	2189731	4

図2) Exist Table 出力結果一部