

生理活性ペプチド同定用シーケンスタグ検索ソフトウェアの開発 Sequence Tag Search Software to Identify Bioactive Peptides

Jingwen Yao · 森本健太郎 · 〇村瀬雅樹 · 船越なつ美 · 福山裕子 · 岩本慎一 · 梶原茂樹 · 田中耕一 (島津製作所 田中最先端研究所)

J. Yao, K. Morimoto, O.M. Murase, F. Funakoshi, Y. Fukuyama, S. Iwamoto, S. Kajihara, K. Tanaka (Koichi Tanaka Laboratory of Advanced Science and Technology, Shimadzu Corp.)



1. Abstract

Identification of bioactive peptides by DB searching is challenging because non-specific processing and PTMs results in very large search space. We developed novel sequence tag searching software to identify peptide variants effectively. For this purpose, focused tag library was built to store sequence tags obtained with high confidence from already identified peptides. As a preliminary test using small set of peptide data, correct sequence tags were derived effectively from undetermined spectra matching with tag library. We also confirmed correct peptides were identified just using top ranked sequence tag.

2. Introduction

2.1 生理活性ペプチドの特徴

- ✓ サイトカインや神経伝達物質等の体内調節因子として生理的に重要
- ✓ 創薬やバイオマーカー開発への応用も期待される
- ✓ 前駆体タンパク質から様々なプロセッシング(切断・修飾)を経て成熟型となる(Fig. 1).
⇒ 様々な切断酵素・修飾により多様な構造をとる
⇒ タンパク質よりもペプチドそのものを同定することが重要
- ✓ 発現量は微量。代謝物も含めて数残基違いのペプチド(免疫化学的手法による判別は困難)も多数生成される。
⇒ 質量分析装置による同定が最適



Fig. 1. Functional/structural variety of endogenous peptide:
For example, pro-opiomelanocortin was cleaved enzymatically into various active peptides.

2.2 従来型のペプチド同定法の問題点

- ✓ 従来の DB 検索法: 切断酵素未知 ⇒ 検索空間増大 ⇒ 同定感度・速度低下
- ✓ *de novo* シーケンス法: 構造多様 ⇒ スペクトル複雑 ⇒ 全長配列推定困難

2.3 新規生理活性ペプチド同定法

以下の仮定に基づき、シーケンスタグ検索法¹⁾の改良法を新規に考案した。本演題では、改良法を用いて実際にペプチド同定が可能か予備検討を行ったので、報告する。

仮定

- ✓ 同定・未同定スペクトルから、シーケンスタグ(3残基長)を推定可能
- ✓ アミノ酸配列を部分的に共有するペプチド群(同一前駆体由来)のうち少なくとも一個を従来法で同定可能
- ✓ 部分配列を共有するペプチドはフラグメントパターンも類似
⇒ 同定スペクトルとの照合により、信頼度の高いタグを導出可能

新規シーケンスタグ検索法 (Fig. 2)

- ✓ 従来法で同定された生理活性ペプチドのスペクトルの特徴を、3残基長のシーケンスタグとタグ構成主要ピークで構成されるタグ情報として抽出し、同定されたペプチドとともにデータベース(タグライブラリ)に登録する。
- ✓ 未同定スペクトル中の主要ピークをタグライブラリ上のタグ構成主要ピークと照合し、シーケンスタグを(*de novo* シーケンシングに比べ)高速に導出。導出したタグを用いてタンパク質 DB を配列検索する(事前の切断サイト・修飾の仮定不要)。

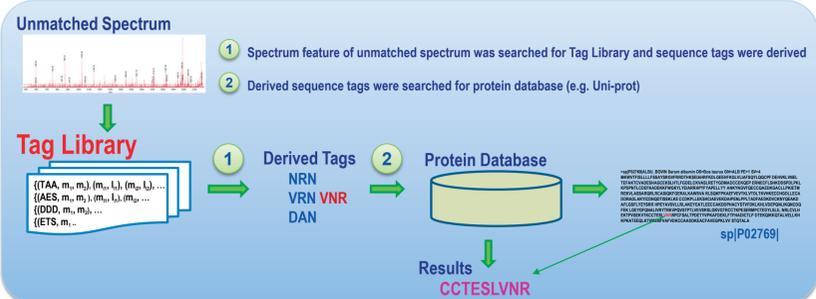


Fig. 2. Novel sequence tag search using focused peptide library

3. Methods

検索アルゴリズム概要(切断サイト・修飾の自動探索)

未同定スペクトルとタグライブラリ上の主要ピークを照合し、マッチしたピークを起点にシーケンスタグ(3残基長)をスコア付きで導出する。スコアの低いタグから順にタンパク質 DB 上の配列を文字列検索し、プリカーサーイオン質量と許容質量誤差に納まるペプチドを探索する (Fig. 3, 4)。

プリカーサー質量に一致するペプチドが見つからなかった場合には、タグ内に当てはまる修飾がないか探索し (Fig. 4a)、合致するものがなければ、タグ外に当てはまる修飾を探索する (Fig. 4b)。

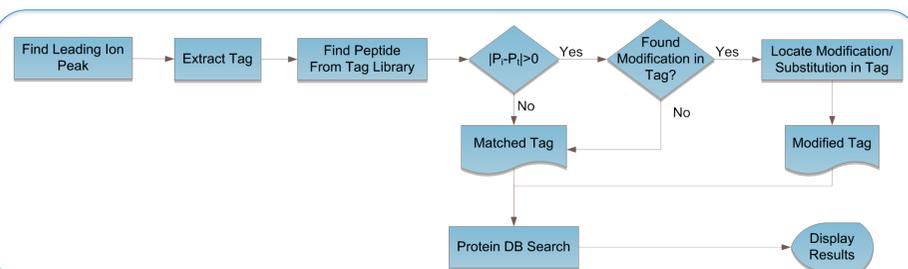


Fig. 3. Flow chart for process from tag query to tag database search:
 P_i : observed peptide ion mass, P_t : theoretical peptide ion mass matched with tag library.

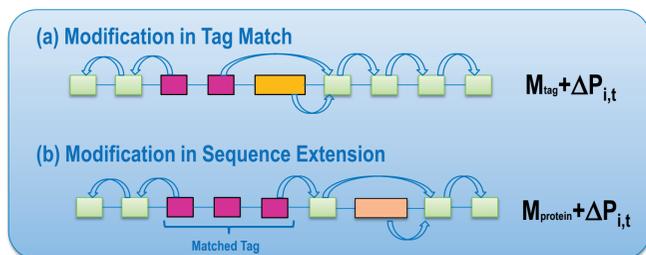


Fig. 4. Schematic diagram of searching for PTM/substitution

4. Results

上記タグ検索アルゴリズムによるペプチド同定が可能か予備的に検証するために、単独の Windows アプリケーションとして検索アルゴリズムを実装した。また、Mass++²⁾から開発ソフトを起動し、ピークリスト作成から検索実行までを円滑に進められるように Mass++ プラグインを開発した。

予備検証として、トリプシン消化物および非トリプシン消化物ペプチドの同定を試みた(テスト1, 2)。タグライブラリには、事前に LC-MALDI MS で測定した BSA トリプシン消化物(5 pmol)の MS² スペクトルから同定された 19 個のペプチドを登録した。タグ導出の動作確認として、同一条件で測定した異なる測定データに対して、開発ソフトを用いて同定を試みたところ、19 例中 15 例について正しいペプチドが 1 位ヒット(低品質でタグの導出ができなかったスペクトル 2 例以外は 3 位までにヒット)した。

テスト1: トリプシン消化物に対する修飾・置換の検索

同定済の β -actin のトリプシン消化物ペプチド (Table 1. No. 1) の MS² スペクトルをタグライブラリに追加登録した。

次に、N 末に formylation を受けた β -actin ペプチド (No. 2)、および、配列が一残基異なるアミノフォーム (α -actin, No. 3) から得られた MS² スペクトルの同定を試みたところ、いずれも 1 位スコアで導出されたシーケンスタグから正しいペプチドが 1 位スコアでヒットした (Fig. 5, 6)。ただし、ペプチド (No. 3) については、他のペプチド由来のタグも同じスコアで導出されており、タグ導出のスコアリングについては改良の余地があると考えられた。

Table 1. Tryptic digest of actin for test 1

No.	Protein	Peptide	Precursor Ion Mass	Mod/Subst
1	β -actin (Human)	QEYDESGPSIVHR	1517 Da.	None
2	β -actin (Human)	QEYDESGPSIVHR	1545 Da.	N-term \rightarrow Formyl [27.8]
3	α -actin (Human)	QEYDEAGPSIVHR	1501 Da.	S \rightarrow A at Position 6

Conclusion

- 生理活性ペプチドの産生・代謝特性を利用して、従来法で同定されたスペクトルの特徴情報を登録したデータベース(タグライブラリ)を活用した新規のシーケンスタグ検索法を考案し、検索ソフトウェアとして実装した。
- トリプシン消化物/非トリプシン消化物の未同定スペクトルをタグライブラリと照合することで、*de novo* シーケンス法を用いずに信頼度の高いシーケンスタグを導出できる可能性が示唆された。
- 事前に切断サイト・修飾を想定せずに、修飾・置換を受けたペプチドを正しく同定できる可能性が示唆された。

Future Work

- 生理活性ペプチドの解析例数を増やし、アルゴリズムの検証・改良を進める。

テスト2: 生理活性ペプチド(非トリプシン消化物)の検索

非トリプシン消化物からもタグの導出および変異の検出が可能か調べた。まず、ヒト由来の obestatin から同定した MS³ スペクトル (Table 2. No.1) をタグライブラリに追加登録した。次に、未同定だった MS² スペクトル (No. 2)、および、ラット由来の obestatin (3 残基違い) の MS² スペクトル (No. 3) に対して検索を試みたところ、1 位スコアで導出されたタグから正しいペプチドが 1 位ヒットした (Fig. 7, 8)。

Table 2. Non-tryptic digest peptide of obestatin for test 2

No.	Sample	Peptide	Precursor Ion Mass
1	obestatin (Human)	VGIKLSGVQYQQHSQAL	1856 Da.
2	obestatin (Human)	FNAPFDVGIKLSGVQYQQHSQAL	2547 Da.
3	obestatin (Rat)	FNAPFDVGIKLSGAQYQQHGRAL	2517 Da.

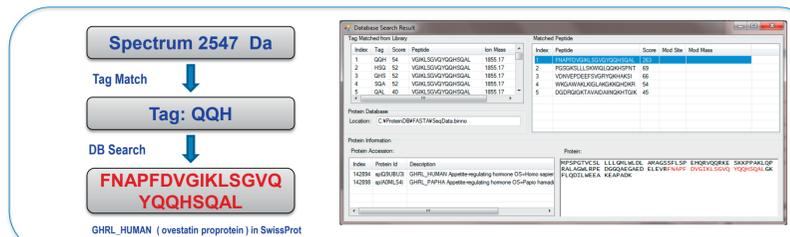


Fig. 7. Search result from spectrum No. 2

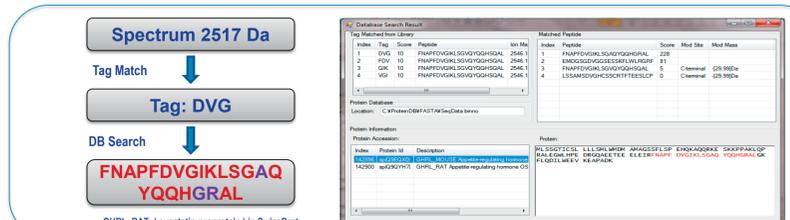


Fig. 8. Search result from spectrum No. 3

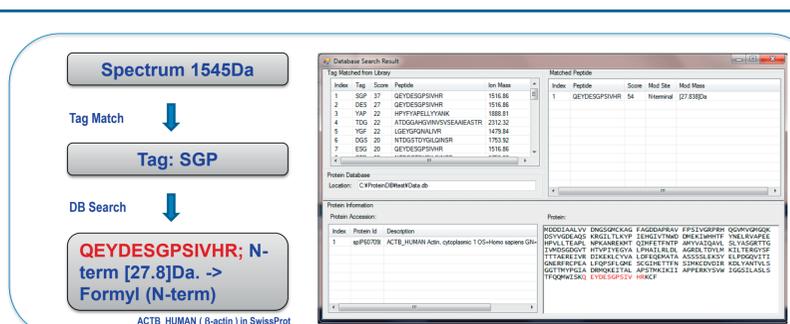


Fig. 5. Search result from spectrum No. 2

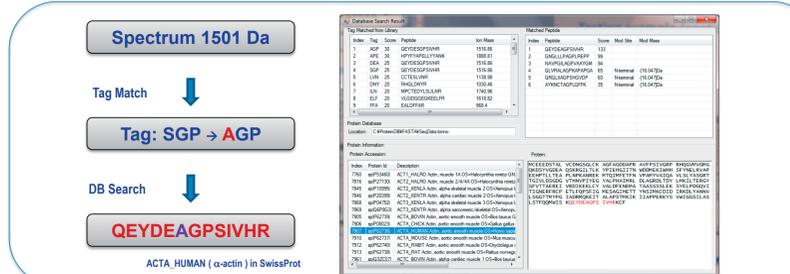


Fig. 6. Search result from spectrum No. 3

Reference

- 1) M. Mann and M. Wilm, *Anal. Chem.*, 1994, 66(24), 4390-4399.
- 2) <http://www.first-ms3d.jp/achievement/software/mass2/>

Acknowledgement

本研究は、日本学術振興会の最先端研究開発支援プログラムにより、助成を受けたものである。