1F1a　質量分析データを用いたゲノム配列へのマッピング

Direct Mapping of Mass Spectrometric Data onto Genomic DNA Sequence

（広島大院理・QuLiS[1]，理研[2]，徳島大分子酵素セ[3]）

○石野洋子[1,2]・岡田ひとみ[2]・谷口寿章[2,3]

質量分析（MS）とデータベース（DB）検索を組み合わせたプロテオーム解析により、細胞や組織で発現している数千種類のタンパク質を同定することが可能となった。そこでは通常、質量分析で得られるペプチドの部分配列情報を用い、タンパク質アミノ酸配列の DB に対して確率統計的な考えに基づく検索を行うことで、試料中のタンパク質の同定を行っている。多くのモデル生物でゲノム DNA 配列が明らかになることで DB の整備が進み、ゲノム DNA とそれらがコードする多くのタンパク質との対応が明らかになることで、この手法が生まれたといえる。しかし、この方法だけでは、DB 上の遺伝子の配列やアノテーションに間違いがある場合に同定精度が落ちる可能性があり、また、まだアノテーションされていない新規の遺伝子は発見できないなどの不十分な面がある。

そこで、本研究では、直接ゲノム DNA の塩基配列 DB に対して確率統計的な検索を行い、検索で得られたペプチドをゲノム配列上にマップすることを試みた。これにより、1) ペプチドの位置と既存のアノテーションとを比較することで配列やアノテーションの間違いを検出する、2) 単純な遺伝子検出プログラムを組み合わせることで新規遺伝子を予測する、という 2 つを可能にした。具体的には、質量分析データの DB 検索エンジンとして Matrix Science 社の Mascot を使用し、統計的な見地からペプチドごとに算出される確からしさのスコアを利用した情報処理を行い、上述の 1)、2)を実施した。

現在までに、枯草菌、分裂酵母等でこの手法を試みているが、今回は、光合成細菌 *Synechocystis* sp. PCC6803 でこの解析を行った結果を示す。異なる光強度下で培養した菌体の可溶性画分と膜画分を試料とし、高速液体クロマトグラフィーを組み込んだ四重極-飛行時間ハイブリッド型質量分析装置（LC-MS/MS）により分析されたデータを、情報解析に用いた。ゲノム配列とアノテーションは GenBank に登録されているものを用いた。その結果、幾つかの新規遺伝子とともにアノテーションの誤りが多く見出されたが、後者の大部分は CyanoBase（http://www.kazusa.or.jp/cyano/）に含まれる新しいアノテーションでは訂正されていた。これらの結果は、プロテオミクスのデータを用いたゲノム解析への応用が有効であることを示している。

（参考文献）

[1] Choudhary, J.S. *et al.*, Matching peptide mass spectra to EST and genomic DNA databases, *Trends in Biotechnology* 2001, 19, S17-22.
[2] Dandekar, T. et al., Re-annotating the Mycoplasma pneumoniae genome sequence: adding value, function and reading frames, Nucleic Acids Res. 2000, 28, 3278-3288.
[3] Himmelreich, R. et al., Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae, Nucleic Acids Res. 1996, 24, 4420-4449.
[4] Jaffe, J.D., Berg, H.C., Church, G.M., Proteogenomic mapping as a complementary method to perform genome annotation, *Proteomics* 2004, 4, 59-77.
[5] Kuster, B., Mortensen, P., Andersen, J.S., Mann, M., Mass spectrometry allows direct identification of proteins in large genomes, *Proteomics* 2001, 1, 641-650.
[6] Link, A.J. et al., Direct analysis of protein complexes using mass spectrometry, *Nature Biotechnology* 1999, 17, 676-682.