PiKA^2: A WWW Based System for Automated Analysis of Protein 2D-PAGE Images

Katsutoshi Takahashi[1] sltaka@jaist.ac.jp
Masayuki Nakazawa[2] nakazawa@kanazawa-tc.ac.jp
Yasuo Watanabe[3] watanabe@infor.kanazawa-it.ac.jp
Akihiko Konagaya[1] kona@jaist.ac.jp

[1] Japan Advanced Institute of Science and Technology (JAIST) Hokuriku, Asahidai 1-1, Tatsunokuchi, Ishikawa 923-1211, Japan
[2] Kanazawa Technical College, Hisayasu 2-270, Kanazawa, Ishikawa 921-8601, Japan
[3] Kanazawa Institute of Technology, Ohgigaoka 7-1, Nonoichi, Ishikawa 921-8501, Japan

Keywords: proteome analysis, automated image processing, protein 2D-PAGE

1 Introduction

The complete sequence of human genomic DNA is predicted to be solved by early in the next century. The powerful genomic technologies now permit the determination of complete genome sequences and exhaustive mRNA expression profiles in cells. In the next stage of genome project, main research topics are being shifted from DNA sequence itself to the biological function of all genes encoded on it and to comprehending the regulation mechanism of complete genetic network. It is insufficient, however, to describe physiological and pathological processes only with the data derived from genetic and genomic techniques. The global analyses of gene expression and gene interaction at the protein level provide a better or complementary basis for the biological processes.

In a past decade, the progress on two-dimensional poly-acrylamide gel electrophoresis (2D-PAGE) techniques has realized the high-throughput, reproducible and high-resolutional separation of over a thousand proteins on a gel. The protein 2D-PAGE techniques have been becoming most important tools for proteome analyses, together with the mass spectrometry (MS) techniques. With the highly-automated equipments, 2D-PAGE gels can be obtained on a massive scale and the proteins resolved as spots can be identified rapidly with MS techniques after automated proteolytic digestion. To promote further automation of high-throughput proteome analyses, the image processing of massive 2D-PAGE gels should also be highly-automated.

In this poster, the computer algorithms which realize fully-automated processing of protein 2D-PAGE images are presented. A World Wide Web (WWW) based system "PiKA^2 (<u>P</u>roteomic <u>I</u>nformation and <u>K</u>nowledge <u>A</u>nalysis for <u>2</u>D-PAGE)", in which the fully-automated algorithms are integrated with the powerful image database, is also presented. With the WWW interface, one can register, search, process and analyze 2D-PAGE images easily, using any networked computer with WWW browser.

2 Computer Algorithms

The fully-automated protein 2D-PAGE image processing algorithms were originally designed for processing 2-DE images of genomic DNA [1, 2, 3], and were modified here for effective protein 2D-PAGE image processing.

The location of the protein spots are detected by the "ring operator" [1] instead of conventional simple peak detector or Gaussian matched filter, in order to avoid ill-recognition of protein spots which have irregular shapes. Once the spot location is detected, the spot shape is modeled by Gaussian-type function [2]. This spot shape modeling realizes not only the accurate spot intensity estimation, but also the identification of the hidden spot, which appears as the shoulder of large protein spot and cannot be detected by any sort of peak detector.

We represented, then, the protein spot pattern on a 2D-PAGE image as a Delaunay net [4], and the equivalent protein spots on two protein 2D-PAGE images are identified by means of simple breadth-first graph search technique [2]. After the pairwise spot pattern matching, spot information on a protein 2D-PAGE image is mapped onto a reference pattern, with correcting non-linear local distortion of the gel [3]. With the only one reference pattern given, multiple protein 2D-PAGE images can be compared all at once. The individual protein spot pattern is matched with the reference pattern and then the protein spots are mapped onto the reference independently. With the hierarchical clustering of the mapped protein spots, the master spot pattern, which unifies the all spot information of multiple 2D-PAGE images at once, can be generated automatically.

With the above algorithms, several thousands of protein spots can be identified and quantified from a gel without any human inspection, and also the spot patterns on several 2D-PAGE gels can be compared rapidly without any human interaction, except for choosing a reference 2D-PAGE image.

3 System Overview

We have integrated the above algorithms as the "PiKA^2" system together with powerful image database and the WWW interface. This system provides multi-user environment and task scheduling mechanisms for 2-DE image processing such as spot identification or spot pattern comparison.

With PiKA^2, massive amount of 2D-PAGE images can be managed easily, and they can be processed and analyzed in batch, to give the objective and fast extraction of proteomic information. In near future, PiKA^2 will be integrated with other databases such as sequence database, EST (expressed sequence tag), PMF (peptide mass fingerprinting) and literature databases, to realize the automated proteomic knowledge management system.

References

[1] Takahashi, K., Nakazawa, M., and Watanabe, Y., DNAinsight: an image processing system for 2–D Gel Electrophoresis of Genomic DNA, *Genome Informatics*, 8:135–146, 1997.

[2] Takahashi, K., Nakazawa, M., Watanabe, Y., and Konagaya, A., Fully-automated spot recognition and matching algorithms for 2–D gel electrophoretogram of genomic DNA, *Genome Informatics*, 9:161–172, 1998.

[3] Takahashi, K., Nakazawa, M., Watanabe, Y., and Konagaya, A., Automated processing of 2–D gel electrophoretograms of genomic DNA for hunting pathogenic DNA molecular changes, *Genome Informatics*, 10:121–132, 1999.

[4] Toussaint, G.T., The relative neighborhood graph of finite planar set, *Pattern Recognition*, 12:261–268, 1980.