

Making a bioinformatics platform for proteome analysis

森澤 拓¹⁾、東方 友久²⁾、山本 伸²⁾、戸田 年総^{1)†}

1) 東京都老人総合研究所プロテオーム共同研究グループ 2) 日本バイオ・ラッドラボラトリーズ[†]

Hiraku Morisawa¹⁾, Tomohisa Toho²⁾, Shin Yamamoto²⁾, Tosifusa Toda^{1)†}

1) Proteomics Collaboration Research Group, Tokyo Metropolitan Institute of Gerontology
2) Nippon Bio-Rad Laboratories[†]

1 はじめに

ゲノム情報の解析には、DNA の塩基配列の特徴を用いて、最新のコンピューターを用いた大規模な相同性検索の手法が大きな役割を果たすことになった。コンピューターの処理速度の向上と低価格化がゲノム解析の研究環境を下支えした。2001 年 2 月にヒトゲノムの草稿が発表されるに至り、ポストゲノム時代が始まることになった。ポストゲノム時代においては、プロテオーム研究による機能解析が重要な位置を占めると考えられる。プロテオーム解析においては、ペプチドマスフィンガープリント (peptide mass fingerprint) 質量分析法で、コンピューターを用いた同一性検索が重要な役割を果たしており、今後も同一性検索において様々な検討が加えられると考えられるものの、プロテオーム情報は、機能解析など、多階層、高次元の情報解析が必要とされており、2次元電気泳動画像などの画像情報も取り込んだ、適切なデータベース化手法の開発が求められている。プロテオーム情報の検索あるいは抽出といった局面において、大量なデータの効率的なデータマイニング手法の検討が求められ、それを可能にするデータベースを実現するために、プロテオーム研究の情報基盤の構築が必要であると考えられる。東京都老人総合研究所では、プロテオーム共同研究センターを設置し、国内外のプロテオーム解析を行う研究者に研究設備を解放すると同時に、システムティックなプロテオーム情報の蓄積と運用のために情報基盤の確立に急いでいる。そこで日本バイオ・ラッド ラボラトリーズ株式会社と共同研究を開始し、プロテオーム解析実験情報の全デジタル化、解析手法のルーチン化、データマイニングのためのプラットフォーム作成、インターネット上でのデータ共有のための XML 化を目指して研究を行った。

2. 方法

100Mbps の研究所内 LAN 上にデータベースサーバ

として、SUN Fire V880 を接続した。OS の Solaris 8 上に Oracle9.2.0.1.0 のインストールを行い、初期設定を行うとともに Bio-Rad 社が開発した WorksBase Software for Proteomics が動作するようデータ構造を作成した。データベースのクライアントパソコンとして、デスクトップパソコン1台、ノートパソコン2台に OS として日本語 Windows2000 Professional と日本語 WindowsXP Home edition 上に Oracle9i Client software および WorksBase software 1.1.2 をインストールした。(Fig 1)

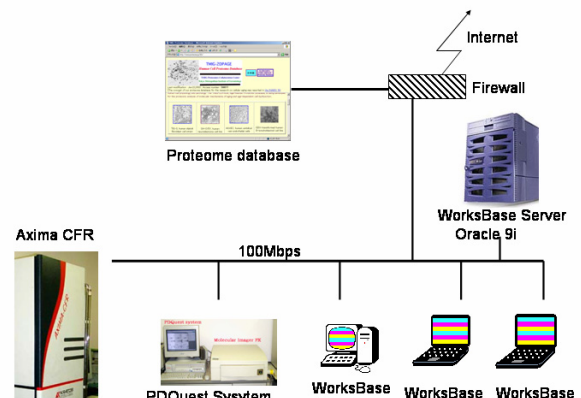


Fig 1 Network for proteome analysis

研究所内 LAN に Bio-Rad 社製 PDQuest 2次元電気泳動画像解析システムと島津製作所製 AXIMA-CFR 質量分析システムを接続した。また、Firewall を介して DMZone にインターネット公開を行う TMIG-2DPAGE (Two-dimensional polyacrylamide gel electrophoresis database) を設置した。

WorksBase software は、実験プロトコル情報の格納は、テキスト形式で保存を行う。レポート形式にテキスト、画像を張り付け、レポートの親子関係を規定する方式になっている。(Fig 2) レポート単位で、ユーザー間の共有、非共有を設定できるようになっていて、実験プロトコルの変遷にしたがい、後から柔軟にセキュリテ

の変更が行える。PDQuest 2次元電気泳動解析システムとデータベースサーバとはオンラインで直結し、画像処理やスポット定量情報などの実験結果の情報は、リアルタイムにデータベースに格納することができるが (Fig3)、Standalone での処理とレスポンスの差があり、一層の高速化が必要である。データベースサーバ上には、インターネット上の公開蛋白質配列データベース SWISS Prot と TrEMBL のデータセットを常に更新しながらダウンロードを行い、実験データとの照合が行える。

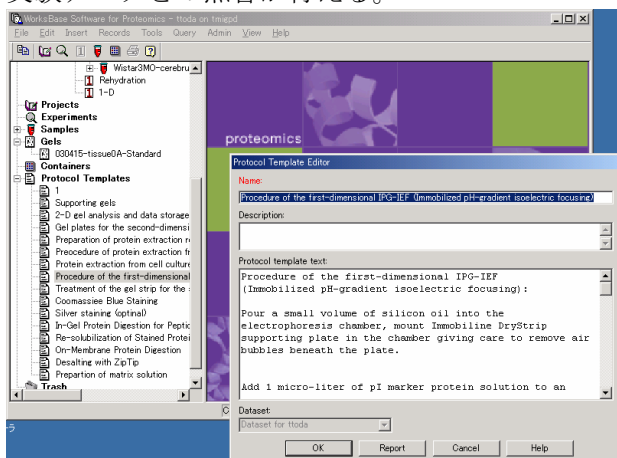


Fig 2. 実験プロトコル等のテキスト入力

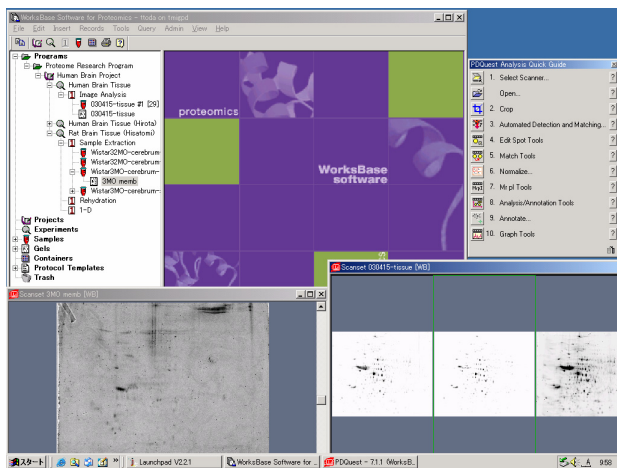


Fig 3. データベースへの画像格納

島津製作所製 AXIMA-CFR 質量分析システムやインターネットに公開している TMIG-2DPAGE とデータベースサーバは、オンラインで接続しているが、手動でデータの移し替えを行った。プロテオームデータの XML フォーマットを作成し、テキストファイルによる XML データベースを作成した。TMIG-2DPAGE のクリックブルマップより XML データが参照できるように設定した。

3. 結果と考察

本データベースサーバは、機器分析より出力

された実験結果などに関しては、検索効率を高め、データマイニングを効果的に行うために、リレーショナルデータベース上でデータベース・スキーマによって整理されたデータベース構築を行い、実験プロトコル等、系統的でない情報に関しては、テキストマイニングの手法が活用できるよう構成されている。実験プロトコルの記述において、結合の仕方は、試行錯誤が必要で、テキストマイニングの検討にあわせ変更する必要がある。今後、系統的でない情報を中心にホームページの検索エンジンのような結合の重み付けにより検索効率をあげていくようなくみを検討したり、またテキストマイニングの手法である、シソーラスに基づいたキーワード抽出やテキスト集合の絞り込みなどの手法の適用が望まれる。さらに、世界的な研究促進を行っていくためにはネットワーク上のデータの共有化に優れた、XML 形式によるデータベース化を視野に入れて柔軟なデータベース構築が必要であると考えられる。WorksBase software は、HTML 形式でのデータの出力をサポートしているため、データ変換ソフトを作成し、プロテオーム XML データベースにルーチンとして接続することが望まれる。より多くの分析機器で出力されたデータをリアルタイムにデータ格納し、機器間でデータ交換を行うシステムに発展させることで、プロテオーム研究を量的な面から下支えし、大きな成果を発揮できると考えられる。

参考文献

1. Taylor CF, Paton NW, Garwood KL, Kirby PD, Stead DA, Yin Z, Deutsch EW, Selway L, Walker J, Riba-Garcia I, Mohammed S, Deery MJ, Howard JA, Dunkley T, Aebersold R, Kell DB, Lilley KS, Roepstorff P, Yates JR 3rd, Brass A, Brown AJ, Cash P, Gaskell SJ, Hubbard SJ, Oliver SG. Related Articles, Links: A systematic approach to modeling, capturing, and disseminating proteomics experimental data. *Nat Biotechnol.* Mar; 21(3):247-54. (2003)
2. Toda T, Kaji K, Kimura N.:TMIG-2DPAGE: a new concept of two-dimensional gel protein database for research on aging. *Electrophoresis.* Feb;19(2):344-8. (1998)
3. 戸田年総:プロテオームデータベースの構築, 現代科学増刊 42 プロテオミクス, pp73-80, 東京化学同人(2002)