

# 遺伝子ネットワークモデル化研究チーム

## Computational Genomics Team

チームリーダー 小長谷明彦

KONAGAYA, Akihiko

遺伝子ネットワークモデル化研究チームでは、生物実験、実験データからの知識抽出、数理モデルによる体系化、シミュレーションによる予測、生物実験による検証というサイクルを繰り返すことにより、細胞内における遺伝子およびタンパク質の相互作用によるダイナミクスの解明を目指している。本年度は、このような研究サイクルを実現するためのソフトウェアプラットフォームとして遺伝子ネットワークシミュレーションと統合データベースの研究に着手した。

### 1. 遺伝子ネットワークシミュレーションの研究 (小長谷, 畠山, 長谷川, 仲<sup>\*1</sup>, 吉岡<sup>\*1</sup>, 川崎<sup>\*2</sup>)

近年、成長ホルモンなどによる細胞内情報伝達機構が分子レベルで急速に解明されているが、各々のタンパク質の関わりは分かっているが、ネットワークレベルでの多種類のタンパク質の相互作用や、量的なダイナミクスに対する理論や研究はここ数年始まったばかりで未知の部分が多い。当研究では連立微分方程式や拡散方程式などを用いたコンピュータシミュレーションと、生体内分子や細胞を使った実験を組み合わせることにより、情報伝達系の鍵となるコンポーネントや反応を見だし、細胞内の情報伝達をシステムの的に理解し、またそれを発がん機構などの解明に役立てることを目的とする。必要な実験系として、培養細胞を使ったモデル、タンパク、ペプチド間等の相互作用、発現プロファイルの多次元的な解析を行う予定である。

#### (1) 遺伝子ネットワークモデル化 (畠山, 仲<sup>\*1</sup>)

細胞内情報伝達系のテストケースとして、上皮増殖因子 (EGF) リセプターの相互作用から MAPK カスケードを経て遺伝子発現に至る系を選択し、数理モデルを作成することにより、必要となる要素技術、ソフトウェア等の確認を行った。反応過程の緩和時間の大きな差異や拡散過程に特徴があることから、必要となる要素技術として、微分方程式によるモデル化とその数値積分によるダイナミクスの解析、モンテカルロ法によるモデル化、さらに大規模な系を表現するものとしてペトリネットがその候補として有効であることが明らかになった。また、数式処理ソフト Mathematica のパターンマッチ・バックトラック機能などを利用したダイナミックな遺伝子ネットワーク・データベースの構築の可能性・有効性についての評価を始めている。今後、細胞周期やさらに転写因子の影響なども加系を拡大し、細胞内情報伝達機構を網羅的にモデル化する予定である。

#### (2) 微分方程式シミュレーションエンジンの開発 (川崎<sup>\*2</sup>)

連立常微分方程式ソルバーおよびその入力インターフェースの開発を行った。化学反応系で広く現れる硬い方程式を解くため、ソルバーの解法アルゴリズムとして、硬い方程式向きのアルゴリズムである数値微分式 (NDF) および二

次の修正 Rosenbrock 式を採用し実装した。入力インターフェースは、Michaelis-Menten 式のものを備え、速度パラメータ等の設定を行う。今後、大規模方程式系を解くことを想定し、ソルバーの並列化を行う。

#### (3) 並列シミュレーション環境の構築 (長谷川)

遺伝子ネットワークシミュレーションのための並列計算機環境設計を行った。主な設計目標は高速な数値シミュレーションを可能にすることと、統合された利用環境を提供することである。高性能並列計算機によるシミュレーションを可能にするため、本システムでは 64 ノード 128 CPU のシミュレーションサーバー式および 8 ノード 16 CPU の開発用サーバー二式において、各ノード間の接続に Myrinet を採用し、さらに SCore をサポートすることによって高性能な並列計算環境を提供することとした。また、本システムに含まれるファイルサーバ、磁気ディスク装置、開発用ワークステーション、Web サーバ等の利用をも含めた透過的なアクセスを可能にするための統合利用環境の提供をも行うこととした。

### 2. 統合データベースの構築 (小長谷, Schoenbach, 長谷川, 吉川<sup>\*3</sup>)

GSC 統合データベースは、大規模非均一かつ分散化されたゲノム情報、トランスクリプトーム情報、プロテオーム情報、フェノーム情報を統合するための柔軟かつ拡張性に富んだネットワーク構造を持ち、日々更新される大量データに対処し、エンドユーザの実用に供することを目的とする。

#### (1) 統合データベースの設計 (Schoenbach)

当研究は、統合データベース委員会の活動の一環として行われ、本年度は現存するゲノム関連データベースならびにソフトウェア製品の調査と評価を行うとともに、データベース構築に必要なデータ形式、システムアーキテクチャ、データベース管理システム、ユーザービュー、セキュリティ、アクセス制御、データ交換形式について設計した。さらに、FANTOM データベース構築に関して、FANTOM 会議に出席し、代替スプライシング部位、繰り返し配列部位などのアノテーションやモチーフ解析を行った。

#### (2) 並列データベース環境の構築 (長谷川)

統合データベース開発・運用のための並列計算機環境の設計を行った。急増するゲノム情報、配列情報、発現プロファイル情報、文献情報等への柔軟かつスケーラブルな対応を考慮した上での高可用性、高信頼性をそなえた情報管理を行える必要がある。そのため、情報管理サーバとして 1 分散システムあたり、64 ビット RISC アーキテクチャプロセッサ 16 個、8 GB 共有メモリ、200 GB ディスク装置の SMP ノードを計 4 ノード搭載した合計 64 プロセッサからなる分散メモリ型クラスタシステムを採用した。各分散

システム間は高速スイッチで接続され、双方向最大1GB/秒の転送速度を有する。さらに、並列ホモロジー検索機能を提供するために解析用サーバの設計を行った。128CPU, 64CPU, 16CPU(2式)を有するPCクラスタシステムで構成し、各ノード間をGigabitイーサネットにて接続することとした。また、並列分散データベースを構築するための、並列通信ライブラリ、分散共有メモリ、グローバルスケジューリングの調査を行った。

(3) 消費者向け遺伝子情報データベースの構築(吉川<sup>\*3</sup>)

ライフサイエンス分野の基礎研究や応用研究には、社会的コンセンサスや臨床家や市民への知識普及と相互理解が不可欠である。今まで日本では、消費者一般や医療人への先端技術に関する情報提供は限られていたが、人組織の利用や人対象試験(治験)の透明化の動きと共に、情報提供と相互コミュニケーションは必須のものとされている。2000年から日本でも情報提供が可能になった、一般向けの治験の被験者募集広告や、急増し始めた医療用医薬品の一般向け情報(広告を含む)の検討、および知識伝達・コミュニケーション齟齬が関係した米国での遺伝子治療による事故のケーススタディーを行った。また、先端分野におけるメディア等の言説特性、認知特性について情報収集と、評価指標の検討のために、インターネットを利用したシステムを開発中である。

\*1 共同研究員, \*2 研究嘱託, \*3 研修生

## 誌 上 発 表 Publications

(原著論文) \*印は査読制度がある論文誌

Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I., Saito T., Okazaki Y., Bono H., Saito R., Kadota K., Matsuo Y., Sakai K., Okido T., Furuno M., Aono H., Carninci P., Sakamoto N., Schönbach C., Suzuki H., Yoshida K., Hayashizaki Y., the RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium: "Functional annotation of a full-length mouse cDNA collection", *Nature* **409**, 685–690 (2001). \*

## 口 頭 発 表 Oral Presentations

(国際会議等)

Kobori T., Maruyama T., Yamaguchi K., and Konagaya A.: "A cellular automata system with reconfigurable hardware: Towards a whole cell simulation", 11th Workshop on Genome Informatics (GIW 2000), (Human Genome Center, University of Tokyo), Tokyo, Dec. (2000).

(国内会議)

小長谷明彦: "ユビキチン-プロテアソーム系による細胞周期制御のモデル化", 第1回バイオシミュレーション研究会・日本人工知能学会第12回分子生物情報研究会, 京田辺, 7月(2000).

小長谷明彦: "分子生物情報科学の展望", 第39回計測自動制御学会学術講演会(SICE2000), 飯塚, 7月(2000).

小長谷明彦: "遺伝子知識スパイラルの構想について", 第13回人工知能学会分子生物情報研究会, 豊中, 9月(2000).

小長谷明彦: "細胞周期シミュレーションに関する一考察", 人工知能学会AIシンポジウム, 東京, 11月(2000).

小長谷明彦: "並列処理はバイオインフォマティクスを救えるか?(パネル討論)", 並列生物情報処理イニシアティブ(IPAB)シンポジウム, 東京, 12月(2000).

小長谷明彦: "バイオインフォマティクスによる遺伝子機能解析", バイオテクノロジー開発技術研究組合シンポジウム「完全長cDNAからの展開」, 東京, 1月(2001).

小長谷明彦: "遺伝子ネットワークモデル化研究に向けて", 第14回人工知能学会分子生物情報研究会, 横浜, 1月(2001).

小長谷明彦: "6N-06 FPGAを用いた遺伝子情報処理のパターンマッチングシステムの設計", 情報処理学会第62回全国大会, 横浜, 3月(2001).

## Research Subjects and Members of Computational Genomics Team

### 1. Computational Genomics

#### Team Leader

Dr. Akihiko KONAGAYA

#### Members

Ms. Mariko HATAKEYAMA

Dr. Christian D. SCHOENBACH

Mr. Aki HASEGAWA

#### Visiting Members

Dr. Takashi HASHIMOTO (Sch. Knowl. Sci., Jpn Adv. Inst. Sci. Technol.)

Dr. Takashi NAKA (Dept. Inf. Sci., Saitama Jr. Coll.)

Mr. Takashi YOSHIOKA (NTT Data Co.)

Mr. Takuji KAWASAKI (Fuji Res. Inst. Co.)

#### Trainees

Ms. Sumi YOSHIKAWA (Sch. Knowl. Sci., Jpn Adv. Inst. Sci. Technol.)