

情報伝達モデル化研究チーム

Computational Proteomics Team

チームリーダー 松 尾 洋
MATSUO, Yo

ゲノムにコードされた情報は、時間・空間特異的に発現されるタンパク質群と核酸・脂質・糖・低分子等との間の相互作用（＝「情報伝達」）のネットワークにより、多種多様な細胞の機能となって実体化される。平成12年4月に発足した当研究チームは、このような生体分子系の機能メカニズムを、タンパク質の立体構造・機能情報に基づき、バイオインフォマティクス・計算生物学・計算化学的手法により解明することを目指している。そのために、理研（ゲノム科学総合研究センター、播磨研究所）による体系的・網羅的・ハイスループットなタンパク質構造・機能決定プロジェクト（RIKEN Structural Genomics/Proteomics Initiative）におけるバイオインフォマティクス・計算生物学研究を分担している。そして、プロジェクトから得られる大量・体系的なタンパク質構造・機能データを有効に活用することにより、タンパク質と他の分子（タンパク質・核酸・脂質・糖・低分子類）との間の相互作用を立体構造に立脚して予測・設計するための研究を進めている。特にこのような分子間相互作用の研究を、創薬・テイラーメイド医療等の課題に活用することを目指している。

1. 情報伝達モデル化研究

（1）構造決定ターゲット選択支援および実験情報・構造決定進捗状況管理システムの開発（松尾，長内，谷，石川*1，田部井*1）

各種生物種のゲノム・cDNA 配列決定プロジェクトの進展に伴い、膨大な数のタンパク質が報告されつつある。その中から、理研の構造ゲノム科学プロジェクトにおけるX線・NMRによる立体構造決定のターゲットとすべきものを、バイオインフォマティクスの手法により選択・優先順位付けしている。

ターゲット候補には、マウス・ヒト・線虫・シロイヌナズナ・高度好熱菌（*Thermus thermophilus* HB8）由来のものを用いている。これら5種由来の約10万件のタンパク質配列データを、公共データベース（NCBI nr等）および理研の配列決定プロジェクト（遺伝子構造・機能研究グループによるマウス完全長cDNA配列セット、理研・阪大による高度好熱菌ゲノム配列）より得た。

ターゲット選択・優先順位付けは、次のような手続きで進めた。BLASTを用いてターゲット候補配列同士を比較した。類似性のカットオフ値として、データベースサイズ 10^8 でEvalue 0.1を用いた。その結果検出されたアミノ酸配列の全体的な類似性に従って、single-linkage clustering法により候補タンパク質をクラスター（ファミリー）に分類した。公共データベース由来の配列に関しては、数残基置換の変異体やフラグメント配列が27%程度（26267/98755配列）含まれていることが分かった。それらの配列の中から

最も長いものを代表として使用することで、全体の計算時間の短縮およびデータ量の圧縮を図った。各ターゲット候補タンパク質に対して、ホモロジー検索・二次構造予測・膜貫通領域予測・コイルドコイル領域予測・機能推定・ホモログの由来生物種の系統パターンの推定、等の様々な情報解析を行った。その結果に基づき、クラスター間に立体構造決定のための優先順位を付けた。これまでのところ優先順位付けの基準として、(i) 後生動物に特異的に分布するもの、(ii) シグナル伝達系、(iii) DNA/RNA 結合（遺伝子転写修復系等）に関係するもの、(iv) 疾患関連、(v) リボソーム・ミトコンドリア局在、を用いた。これらの基準による優先順位付けに基づき、構造決定のターゲットを選択することになる。なお、配列情報解析の結果、構造既知のホモログが検出されたものや、水溶性球状でないかと推定されるタンパク質（ドメイン）の優先度は低くした。以上の情報解析結果に基づき選択されたターゲットは、クローニング・大量発現・可溶化・精製・結晶化などのステップを経て、X線結晶解析・NMRにより立体構造が決定される。

上記の情報解析の過程、およびその後の立体構造決定の過程で得られる様々なデータを組織的に集積した総合的データベースシステムを開発している。当システムは、配列情報解析・優先順位付けの結果をグラフィカルユーザインターフェイスを通じて提示する。それにより、ユーザの判断に応じた最終的なターゲット選択・優先順位付けを支援する。一方、選択されたターゲットの立体構造決定に至る過程には、プライマー作成・ベクター作成・大量発現・精製・NMR測定あるいは結晶化・結晶解析等のような様々な段階がある。これら各段階に関するデータが収集され、当システムの一部をなす構造決定進捗状況管理システムに格納される。ユーザは、選択された多数のターゲットタンパク質それぞれの構造決定がどの段階まで進んでいるかを一覧することができる。以上のようなシステムを、RDBMSとWWW・サーバーサイドスクリプト技術の使用により実装した。当データベースシステムは、ユーザ認証・IPフィルタリング等によりセキュリティを確保した上で、プロジェクト内部のメンバーの使用に供されている。また、データベースの内容の一部は、<http://www.rsg.riken.go.jp/>を通じて外部に公開されている。

（2）進化的・機能的観点からのタンパク質分類・体系化（松尾，谷，佐藤）

構造ゲノム科学プロジェクト推進のためのターゲット候補タンパク質のクラスター分類に関連して、進化的観点からの既知タンパク質群の体系的階層分類をアミノ酸配列・立体構造比較に基づき進めている。進化的観点を厳密に考慮した階層分類は、構造決定ターゲットの効率的選択、つまり全タンパク質ファミリーの効率的・効果的網羅のため

に実用的な意義がある。同時に、現存の個別タンパク質が持つ生理的機能の由来と生体システムの中での役割を理解するために有用である。

そのような進化的分類を行うためにまず、タンパク質のドメインをアミノ酸配列から推定するための方法を開発した。タンパク質（遺伝子）の進化の単位はドメインと考えられるからである。また、NMRによる立体構造決定はドメイン単位で行う必要があるため、当方法による推定はNMRによる構造決定を進める上でも有用である。当方法では、あるタンパク質のドメインを同定するために、そのアミノ酸配列を用いて既知タンパク質配列データベースを検索する。そしてそのアミノ酸配列上で、既知タンパク質との間に有意な類似性を示した一群の局所領域を収集する。こうして見いだされた局所領域同士を比較し、領域としての重なりが大きいものは1つの領域へとまとめていく。その結果得られる領域をドメインと推定する。当方法を、NMRによる立体構造決定の現時点での主要なターゲット候補となっているマウス由来タンパク質セットに対して適用した。推定されたドメイン群のアミノ酸配列同士を比較し、優位な類似性を示すものをクラスター（ファミリー）にまとめた。これにより、ドメイン単位での構造決定ターゲットの選択・優先順位付けを行うことが容易になった。

さて、一般に進化的関係が遠くなるとタンパク質間のアミノ酸配列の類似性は段々低くなり、ついには有意な配列間類似性を検出することは困難になる。そのように遠い進化的関係にあるタンパク質同士であっても、主鎖骨格の基本的立体構造パターン（フォールド）はよく保存されることが経験的に知られている。このことを利用して、アミノ酸配列比較では困難な遠い進化的関係の検出を立体構造比較により行うことが、体系的なタンパク質分類体系の構築のために必要となる。タンパク質立体構造の類似性の検出のための方法はこれまでに多数提案されている。そこで、既存の諸方法を組み合わせた上で細部に改良を施すことにより、当研究チームの研究目的のために最適化された方法を開発している。特に、タンパク質立体構造データベース検索等の大量計算に耐え得るだけの高速性を持ちつつ、既存の方法と較べて精度を落とさないための工夫を重ねている。当方法は、構造データベースの高速な検索を可能にするために、主鎖骨格構造データを文字列に変換してそれら文字列同士を比較することにより、類似構造領域を持つ可能性のある構造対のみを選別する。一対一の構造比較においては、まずC α 原子間距離行列の比較により数残基長のアラインメントのコアを検出し、それに基づき反復的に立体構造同士を重ね合わせながら類似構造領域全体を検出する。現在、精度・計算時間に関して他の方法との比較を行いながら、細部の修正を重ねている。

（3）生体分子間相互作用の予測・設計（松尾，Pichierri，長内，坂居，大野^{*2}，白川^{*2}）

個々のタンパク質の立体構造データは、それらタンパク質群が生体分子系の中でどのような働きを協調して行うかを理解するための基盤を与える。そこで、タンパク質立体構造情報に立脚して、タンパク質と他の分子との相互作用（＝分子機能）に関する研究を、データベースを利用したknowledge-basedの情報解析とエネルギー計算・分子力学・自由エネルギー摂動法などを用いた分子シミュレーション

の手法を用いて行っている。これらの研究は、薬剤候補化合物の*in silico*スクリーニング・設計等にも役立つであろう。一方、SNPsを中心とするDNA多型が大量に収集されつつある。同時に、DNA多型と疾患感受性・薬剤応答性等の表現型の個体差との相関に関するデータも蓄積されている。遺伝子型を個体レベルの表現型へ結び付けるのは、遺伝子産物であるタンパク質を中心とする生体分子系である。そこで、特にアミノ酸置換をもたらすcSNPsと転写制御領域中のSNPsに注目し、これらSNPsがタンパク質機能・タンパク質-DNA相互作用の変化を通じて生体分子系に与える影響を予測することを目指している。これらの研究は、体質に応じた療法・薬剤の合理的選択や薬剤の合理的設計（テーラーメイド医療）に役立つであろう。

本年度は、以上の研究を今後展開するための基盤を作るために、いくつかの具体例を選んで情報解析・分子シミュレーション手法を適用し、それら手法の有効性を検証した。具体例として、選択的細胞内輸送において重要な役割を担うAP複合体の成分である μ タンパク質ファミリーと、アトピー・喘息等のアレルギー疾患の発症に関わるIL-13・IL-13レセプターを選んだ。

μ タンパク質ファミリーは、被覆小胞に局在するAP複合体の成分として、輸送対象となるタンパク質の細胞質側ペプチドを認識する。認識されるペプチド配列には、XXXXXX ϕ （ ϕ は疎水性残基）というパターンがあることが知られている。このパターンを持つペプチドと μ タンパク質との相互作用の有無・強弱が、金沢大学・大野博司教授らにより酵母2-ハイブリッド法を用いて体系的に調べられている。また、他の研究グループからも、いくつかのペプチドに対する解離定数が報告されている。そこでこれらの既知データを用いて、ペプチドと μ タンパク質との結合の強弱を計算機シミュレーションで予測できるかどうかを確認することにした。 μ タンパク質ファミリーには、特定の細胞・組織特異的発現を示すいくつかのメンバーが知られている。今回の研究では、 $\mu 1 \cdot \mu 2$ （立体構造既知）・ $\mu 3A \cdot \mu 3B \cdot \mu 4$ を対象にしている。まず、 $\mu 2$ の既知構造をテンプレートにしたホモロジーモデリングにより、他の μ タンパク質の立体構造モデルを作成している。一方、 $\mu 2$ とペプチドASDYQRL・AADYQRLとの結合の自由エネルギー変化を、自由エネルギー摂動法およびthermodynamic integration法とにより評価している。予備的な計算を実験値と比較した結果、結合の強弱に関する定性的な傾向を再現することができた。今後は、さらに多くのペプチドに対して適用することにより、結合自由エネルギー変化をどの程度定量的に予測できるかを評価する予定である。

IL-13の関与するシグナル伝達系はアトピー・喘息等のアレルギー疾患感受性に関与することが知られている。そして、IL-13、IL-13レセプター（IL-4R α とIL-13R $\alpha 1$ のヘテロダイマー）に関して、これら疾患への感受性と相関するSNPs（single nucleotide polymorphisms）が、白川太郎教授（京大）らによりいくつか報告されている。これらSNPsのあるものは、IL-13のIL-13レセプターへの結合の強さに影響を与えると推定されている。そこで、このことを計算機シミュレーションにより確認することにした。まず、gp130、ヒト成長ホルモン、ヒト成長ホルモン：レセプター複合体、IL-4R α 等のホモログの既知構造をテンプレ

レートにして, IL-13, IL-13R α 1, IL-4R α :IL-13R α 1 複合体, IL-13:IL-13R 複合体のモデル構造を, ホモロジーモデリングにより作成している。

以上のような計算機シミュレーションに加えて, 分子間相互作用の解析・設計をグラフィックスや仮想現実 (Virtual Reality) 技術を用いて支援するために, 分子設計支援システムを作成している。このようなシステムを独自に開発することは, 次のような理由から必要である。つまり, (i) 当研究チーム特有の研究を効果的に支援するため, (ii) 当研究チームで行っている他の研究と効率的に連携するため, また, (iii) 当研究チームで開発したソフトウェア・データベース群との効率的な連携・統合を図るため, 等である。これまでのところ当システムのうち, 分子構造の表示モジュール (様々な表示法が可能), 同一画面上に複数分子を表示し独立に操作するためのモジュール, 市販の VR デバイス PHANTOM (三次元ポインティング・触覚デバイス) へのインターフェイスとそれによる操作のためのモジュール, を作成した。今後は, 単一シーンに対する複数画面の同時表示, ネットワーク経由での操作・コミュニケーション機能, 分子の変形・改変・設計のためのモジュール, 分子動力学の力場に基づく触感の実現, 等を進める予定である。

*1 業務委託, *2 共同研究員

誌上発表 Publications

(原著論文) *印は査読制度がある論文誌

Nakatsu F., Sakuma M., Matsuo Y., Aarase H., Yamasaki S., Nakamura N., Saito T., and Ohno H.: “A di-leucine signal in the ubiquitin moiety: Possible involvement in ubiquitination-mediated endocytosis”, *J. Biol. Chem.* **275**, 26213–26219 (2000). *

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). *

Kuroda Y., Tani K., Matsuo Y., and Yokoyama S.: “Automated search of natively folded protein fragments for high-throughput structure determination in structural genomics”, *Protein Sci.* **9**, 2313–2321 (2000). *

(総説)

Yokoyama S., Hirota H., Kigawa T., Yabuki T., Shirouzu M., Terada T., Ito Y., Matsuo Y., Kuroda Y., Nishimura Y., Kyogoku Y., Miki K., Masui R., and Kuramitsu S.: “Structural genomics projects in Japan”, *Nat. Struct. Biol.* **7** Suppl., 943–945 (2000).

Yokoyama S., Matsuo Y., Hirota H., Kigawa T., Shirouzu M., Kuroda Y., Kurumizaka H., Kawaguchi S., Ito Y., Shibata T., Kainosho M., Nishimura Y., Inoue Y., and

Kuramitsu S.: “Structural genomics projects in Japan”, *Prog. Biophys. Mol. Biol.* **73**, 363–376 (2000).

松尾洋: “構造ジェノミクスとプロテオミクス”, *遺伝子医学* **4**, 568–572 (2000).

松尾洋: “バイオインフォマティクスからの創薬”, *医学のあゆみ* **196**, 415–418 (2001).

松尾洋: “タンパク質の設計原理を解く”, *細胞工学* **20**, 74–78 (2001).

松尾洋: “構造決定ターゲットの選択と優先順位付け”, *生物物理* **40**, 305–306 (2000).

口頭発表 Oral Presentations

(国際会議等)

Chikayama E., Kuroda Y., Tani K., Matsuo Y., Kawai J., Hayashizaki Y., and Yokoyama S.: “Automated protein domain prediction system for the selection of structure determination targets”, *Int. Conf. on Structural Genomics 2000 (ICSG2000)*, Yokohama, Nov. (2000).

Kuroda Y., Tani K., Matsuo Y., and Yokoyama S.: “Automated search of natively folded protein fragments for high-throughput structure determination in structural genomics”, *Int. Conf. on Structural Genomics 2000 (ICSG2000)*, Yokohama, Nov. (2000).

Tani K., Osanai T., Matsuo Y., and Yokoyama S.: “Development of a system for selection and prioritization of targets for structure determination in structural genomics”, *Int. Conf. on Structural Genomics 2000 (ICSG2000)*, Yokohama, Nov. (2000).

Eda K., Kurumizaka H., Kagawa W., Koshiba S., Kigawa T., Matsuo Y., Nojima H., and Yokoyama S.: “Domain analysis of the human Mcm6 protein”, *Int. Conf. on Structural Genomics 2000 (ICSG2000)*, Yokohama, Nov. (2000).

Pichierri F. and Matsuo Y.: “Homology modeling of interleukin-13 and its genetic variants”, *Computational Science Workshop 2001, (JRCAT, RICS)*, Tsukuba, Mar. (2001).

(国内会議)

松尾洋: “構造ジェノミクスとゲノム創薬”, 第10回コンビナトリアルケミストリー研究会, 大阪, 4月 (2000).

松尾洋: “構造ジェノミクスにおけるバイオインフォマティクス: 蛋白質構造・機能情報に基づく生体分子系機能メカニズムの解明”, 「高次ゲノム科学」研究会, 三島, 7月 (2000).

松尾洋: “構造ジェノミクスとバイオインフォマティクスの現状と創薬への活用”, 技術情報協会セミナー, 東京, 9月 (2000).

谷一寿, 長内隆, 黒田裕, 松尾洋, 横山茂之: “構造ゲノミクスにおける立体構造決定ターゲット選択支援のためのシステム開発”, 第38回日本生物物理学会年会, 仙台, 9月 (2000).

近山英輔, 柳在圭, 黒田裕, 松尾洋, 谷一寿, 河合純, 林崎良英, 横山茂之: “構造ゲノム科学における構造解析の対象となるタンパク質ドメイン領域の自動推定システム”, 第38回日本生物物理学会回年会, 仙台, 9月 (2000).

黒田裕, 松尾洋, 谷一寿, 横山茂之: “相同配列解析に基づく自動的なドメイン推定”, 第 38 日本生物物理学会回年会, 仙台, 9 月 (2000).

松尾洋: “グルタミン酸受容体について知りたい”, 「生物物理バイオインフォマティクス講習会: 配列から立体構造, ドッキングまで」, (日本生物物理学会), 東京, 11 月 (2000).

松尾洋: “構造ジェノミクスにおける構造・機能分類”, 大阪大学蛋白質研究所セミナー「タンパク質立体構造の分類・予測・デザイン」, 大阪, 12 月 (2000).

矢吹孝, 元田容子, 松田夏子, 黒田裕, 松尾洋, 林崎良英, 木川隆則, 横山茂之: “PCR と無細胞タンパク質合成系を用いた, 多検体, 高収量, 迅速なタンパク質ドメインの発現法”, 第 23 回日本分子生物学会年会, 神戸, 12 月 (2000).

谷一寿, 白水美香子, 胡桃坂仁志, 長内隆, 松尾洋, 横山茂之: “構造ジェノミクスにおける構造決定ターゲットの選択と優先順位付け”, 第 23 回日本分子生物学会年会, 神戸, 12 月 (2000).

長内隆, 谷一寿, 松尾洋, 横山茂之: “構造ジェノミクスのためのタンパク質情報解析・実験情報管理総合システム”, 第 23 回日本分子生物学会年会, 神戸, 12 月 (2000).

松尾洋: “構造ジェノミクスの進展”, 第 23 回日本分子生物学会年会, 神戸, 12 月 (2000).

松尾洋: “構造ジェノミクスの展開”, 第 14 回人工知能学会分子生物情報研究会, 横浜, 1 月 (2001).

Research Subjects and Members of Computational Proteomics Team

1. Computational Proteomics

Team Leader

Dr. Yo MATSUO

Members

Dr. Fabio PICHIERRI

Mr. Takashi OSANAI

Mr. Kazutoshi TANI

Dr. Kazuhito SATOU

Mr. Saburoh SAKAI

in collaboration with

Mr. Mitsuyoshi ISHIKAWA (DYNACOM Co., Ltd.)

Mr. Norio Tabei (DYNACOM Co., Ltd.)

Visiting Members

Dr. Hiroshi OHNO (Kanazawa Univ.)

Dr. Taro SHIRAKAWA (Grad. Sch. Pub. Health, Kyoto Univ.)

Dr. Jill E. GREADY (Australian Natl. Univ., Australia)

Dr. Stephen H. BRYANT (NCBI, NLM, NIH, USA)

Dr. Steven E. BRENNER (Univ. California, Berkeley, USA)

Dr. Janet M. THORNTON (Univ. College London, UK)

Trainees

Ms. Akiko FUJINO (Fac. Sci., Univ. Tokyo)

Mr. LIU Zheng (Fac. Sci., Univ. Tokyo)

Mr. Mario ALBRECHT (Univ. Bonn, Germany)